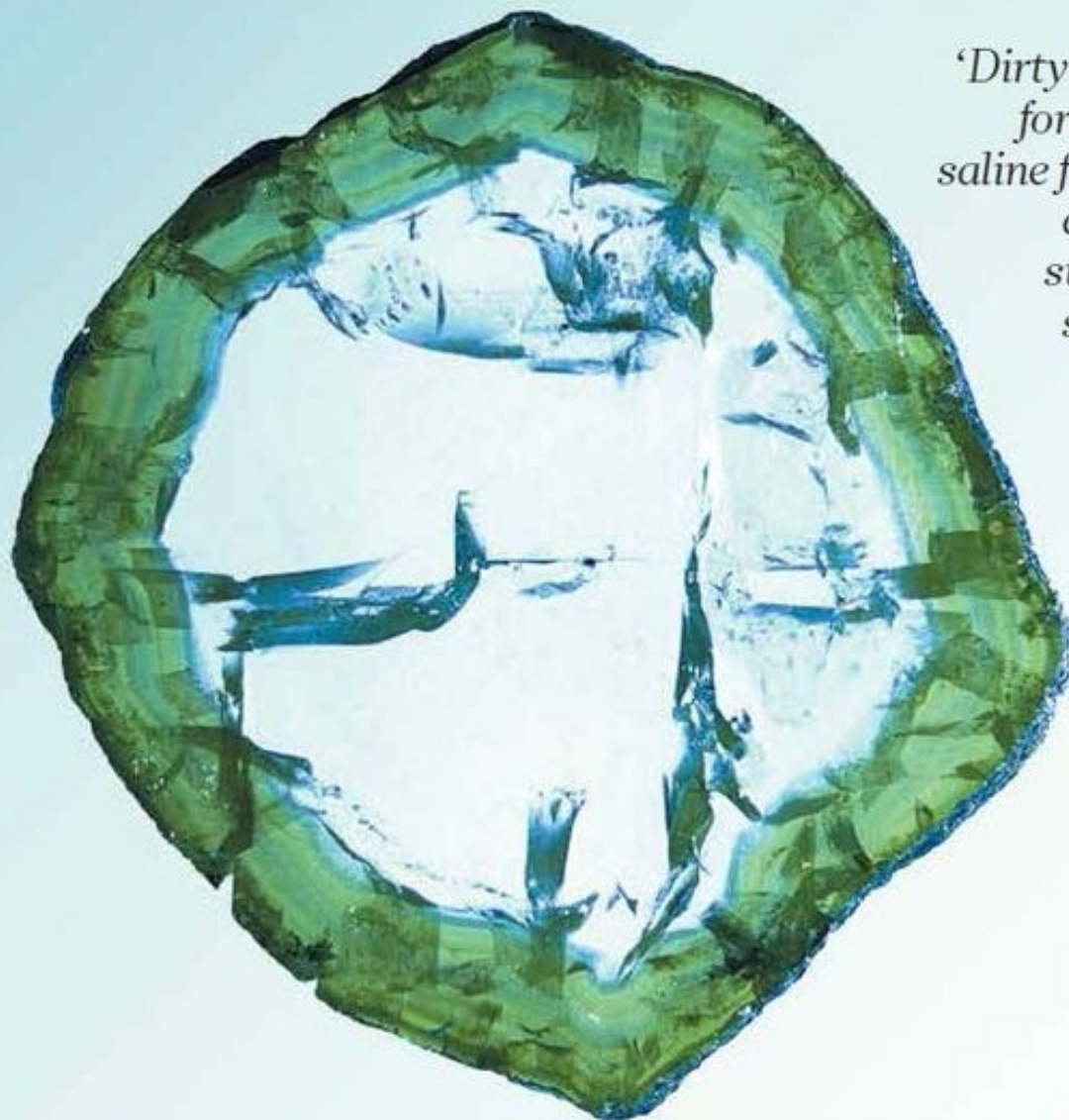


nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



'Dirty' diamond
formation by
saline fluids from
an ancient
subducting
slab **PAGE 339**

A PINCH OF SALT

PUBLIC HEALTH

THE CANNABIS EXPERIMENT

*Relaxed laws fuel a new
science of marijuana use*

PAGE 280



CLIMATE ADAPTATION

15 COASTS TO DEFEND

*Global priorities
10 years after Katrina*

PAGE 285

SOLAR SYSTEM

PLANETARY GROWTH

*Small pebbles are key
to gas giant formation*

PAGE 322

NATURE.COM/NATURE

20 August 2015 £10

Vol. 524, No. 7565



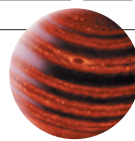
9 770028 083095

THIS WEEK

EDITORIALS

WORLD VIEW Vaccination programme urgently needed to tackle typhoid in Nepal **p.267**

ASTRONOMY Gas giant is lowest-mass exoplanet imaged directly **p.268**



NEUROSCIENCE Dreams spark visual activity in the brain **p.269**

Rise of the citizen scientist

From the oceans to the soil, technology is changing the part that amateurs can play in research. But this greater involvement raises concerns that must be addressed.

Science is not just for scientists these days. Going on a scuba-diving holiday this summer? Share the temperature data from your dive computer with researchers eager to plug holes in sparse records for inshore areas. Nervous about possible pollution from a nearby fracking project? Ease your concerns by helping to collect and analyse air samples as part of a monitoring project. Stuck at home as the rain pours down? Log on to the Internet and spend a couple of hours folding proteins and RNA to help university scientists work out how biology does it.

Citizen science has come a long way from the first distributed-computing projects that hoovered up spare processing power on home computers to perform calculations or search for alien signals. And it has progressed further still since the earliest public surveys of wildlife: it was way back in 1900 that the Audubon Society persuaded Americans to exchange their Christmas tradition of shooting birds for a more productive effort to count them instead.

Some professional scientists are sniffy about the role of amateurs, but as an increasing number of academic papers makes clear, the results can be valuable and can help both to generate data and to inform policy.

A paper in *Geoderma* entitled ‘Can citizen science assist digital soil mapping?’ (D. G. Rossiter *et al.* *Geoderma* **259–260**, 71–80; 2015) makes the case that, yes, non-specialists can help expert soil scientists to track quality, properties and types of soil. It goes further: these amateur soil researchers should be recruited to help with existing and future national surveys. Civil engineers and construction workers routinely view the subsoil, and digging foundations for buildings and trenches for pipelines offers a unique look at the spatial variability of different layers. An army of geocachers — twenty-first-century treasure hunters — visit harsh terrain and difficult-to-access places, and could collect soil data. And they routinely use satellite navigation to record their journeys.

Technology can make scientists of us all. Data churned out by the rapid spread of consumer gadgets equipped with satellite navigation, cameras and a suite of other sensors, and the ease of sharing the results digitally, are driving the boom in citizen science. Volunteers can already identify whale songs from recordings, report litter and invasive species, and send in the skeletons of fish they have caught and consumed. But there is more to being a scientist, of course, than collecting and sharing data — especially if the results are to be used to help determine policy.

Critics have raised concerns about data quality, and some studies do find that volunteers are less able to identify plant species than are academics and land managers. And there are issues around how to reward and recognize the contribution of volunteers, and around ensuring that data are shared or kept confidential as appropriate. But these problems seem relatively simple to address — not least because they reflect points — from authorship to data quality and access — that the professional scientific community is already wrestling with.

More troubling, perhaps, is the potential for conflicts of interest. One reason that some citizen scientists volunteer is to advance their political objectives. Opponents of fracking, for example, might help to track possible pollution because they want to gather evidence of harmful effects. When Australian scientists asked people who had volunteered to monitor koala populations how the animals should be managed, they found that the citizen scientists had strong views on protection that did not reflect broader public opinion.

Scientists and funders are right to encourage the shift from passive citizen science — number crunching — to more-active roles, including sample collection. But as increased scrutiny falls on the reliability of the work of professional scientists, full transparency about the motives and ambitions of amateurs is essential. ■

Portfolio boost

People who fund or manage many research projects could do it better with mindful analysis.

Imagine that you are in charge of a foundation funding a substantial range of research projects, or a university wanting to understand the broader benefits of its research. With a clear view of your institution's overarching mission, you seek good researchers. You might then conclude: as long as they deliver impactful research, your job is done.

But you could benefit from standing back and analysing the situation through different lenses. For example, you could check whether research projects display undesirable uniformity in their underlying assumptions. You could enable collaboration between research strands

that happen to be converging on a particular social or technological outcome. Or you might find ways to diversify, maximizing the chances that something valuable (such as a diagnostic) will emerge even if the central goal (such as a vaccine) turns out not to be achievable.

Such are the virtues of research-portfolio analysis. Those intrigued by such mindfulness should turn to a study (M. L. Wallace and I. Rafols *Minerva* **53**, 89–115; 2015) that looks at relevant literature from recent decades, and highlights the benefits of inward-looking analysis that also creates opportunities for transparency and stakeholder engagement. The study highlights the limitations of corporate approaches to portfolio analysis that are geared to financial returns and well-defined markets — in contrast to the broader aims of government, philanthropic funders and universities.

In analysing how projects and themes can be mapped in ways that highlight “cognitive proximity”, this work may help those who run research portfolios to get better bangs for their bucks. It might even help to make the world a better place. ■



Tackle Nepal's typhoid problem now

As post-earthquake conditions increase the risk of a typhoid epidemic, Buddha Basnyat calls for a widespread vaccination programme.

The news that a promising vaccine has been tested against Ebola is very welcome. But, as we in Nepal know only too well, the development of a vaccine is no guarantee that it will be used. Back in 1987, a successful trial of a typhoid vaccine here in Kathmandu gave the world a new and highly effective way to protect against one of its oldest killers. Yet, ironically, while the tourists and trekkers who have flocked in recent decades to Nepal — one of the countries where typhoid is still endemic — use the vaccine to protect themselves, most local people are denied it.

The Nepalese government cannot afford the US\$10 doses of vaccine that could protect millions of its people. But the need is great — particularly since the devastating magnitude-7.8 earthquake that rocked Nepal in April. In many places, whatever water and sewerage systems existed were smashed, and thousands of people still live in temporary shelters and camps, increasing their risk of consuming contaminated water. In recent months, the monsoon (when cases of typhoid typically peak) has made a bad situation worse, and there are sporadic reports of disease outbreaks. These could yet provoke an epidemic in the hardest-hit places, such as the Sindhupalchowk and Gorkha districts.

As typhoid spreads, the risks increase. The most severely affected are school-age children, who are struck down with a fever and abdominal pain for several weeks. Most of these children recover, but many develop complications, including hypotensive shock, perforation of the gut and gastrointestinal haemorrhage. Antibiotic treatment has reduced the fatality rate from historic highs of 20% to 1–4%, but studies suggest that some typhoid strains are becoming resistant. Antibiotics are sold freely over the counter in Nepal and so are widely misused. About one in three people who report to a Kathmandu doctor with suspected typhoid will already have (mis)used an antibiotic.

The aftermath of the earthquake has made many poor people even poorer. With no health insurance, the average Nepali family now has difficulty paying for an ill child to go to hospital. As a result, those children who develop typhoid complications are less likely to recover.

Prevention is key. The World Health Organization (WHO) said as much in 2008, when it recommended programmes to vaccinate school-age children against typhoid. A year later, the WHO South-East Asia Regional Office recommended that such programmes be prioritized for “immediate” implementation. Yet, despite several promises and false starts, this has not happened in Nepal.

The best way to prevent typhoid is to provide access to clean water and sanitation. But it will take decades for Nepal to put the necessary infrastructure in place, and the impact of the earthquake will make it harder still. In the

meantime, wide deployment of a vaccine could be a useful — and lifesaving — measure. As part of the post-earthquake recovery plan, Nepal should kick-start a long-promised vaccination programme.

Part of the problem is that the typhoid vaccine recommended by the WHO — the one that was tested in Kathmandu — is among the most expensive. To buy and deliver enough doses, Nepal will almost certainly need financial support. This could come from Gavi, the Vaccine Alliance, based in Geneva, Switzerland, which is committed to increasing access to vaccination in poor countries, or from the profits of pharmaceutical companies. (Incidentally, another useful vaccine, against hepatitis E — a disease with a similar mode of transmission that can prove fatal to pregnant women — was also proven in Kathmandu. It, too, is unavailable to local people.)

At present, just one typhoid vaccine is approved by the WHO. It would be cheaper for Nepali health officials to deploy an alternative vaccine, not pre-qualified by the WHO. It is unclear why they have not yet done so; perhaps they are reluctant to be seen to be going against WHO guidance.

The decision to vaccinate — whichever vaccine is used — must be backed with solid research. Here, there is some progress to report. Early next year, a multi-site project led by researchers at the University of Oxford, UK, will start to map the burden of typhoid fever across Asia and Africa. The objective of this study is to quantify accurately the number of people with typhoid fever and complications of the disease in countries such as Nepal. This information will be useful for vaccination campaigns.

There are other pressing research needs, too. With no new treatments in the pipeline, it is important to track resistance to antibiotics. This can be done with proper microbiological support, including, where possible, genetic studies of the typhoid isolates.

Researchers must also study why some people are more susceptible to the disease more than others. Last year, a preliminary genome-wide association study of people from Nepal and Vietnam implicated the *HLA-DRB1* gene as a major contributor to resistance against typhoid fever (S. J. Dunstan *et al.* *Nature Genet.* **46**, 1333–1336; 2014).

If the Ebola vaccine is as effective as it seems to be, the world will have developed a way to tame this awful disease within 40 years of its first appearance. Reports of typhoid date back almost 2,000 years, but the disease is still killing people. It is not just the new threats that deserve our attention. ■

Buddha Basnyat practises medicine in Kathmandu, where he is director of the Oxford University Clinical Research Unit at Patan Hospital.
e-mail: buddha.basnyat@ndm.ox.ac.uk

WHILE TOURISTS AND TREKKERS USE THE VACCINE TO PROTECT THEMSELVES, MOST LOCAL PEOPLE ARE DENIED IT.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/hvbykk

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

MOLECULAR EVOLUTION

How yeast got a second genome

The origin of the duplicate genome in baker's yeast has been revealed: it arose when cells of two different species fused early in yeast's evolution.

Whole-genome duplication has happened during the evolution of many organisms, including yeast — in which researchers had thought it was a spontaneous event.

Marina Marcet-Houben and Toni Gabaldón at the Centre for Genomic Regulation in Barcelona, Spain, analysed the genomes of *Saccharomyces cerevisiae* to reconstruct the evolutionary history of each of the yeast's genes. They concluded that each of the two subgenomes originated from a different lineage. Over time, there were extensive genomic rearrangements and gene loss.

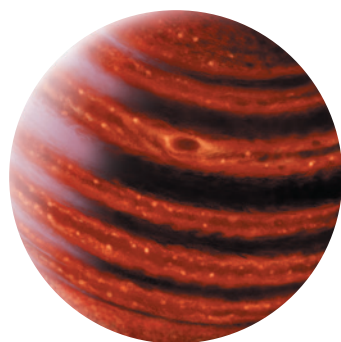
This hybridization could have given yeast an advantage by bringing together different physiological properties.

PLoS Biol. 13, e1002220 (2015)

ASTRONOMY

Direct look at a small exoplanet

Astronomers have snapped a picture of a planet like those in the Solar System but orbiting another star — the lowest-mass exoplanet ever directly imaged.



Bruce Macintosh of Stanford University in California and his team used the Gemini Planet Imager at the Gemini South Telescope in Chile to photograph the planet orbiting 51 Eridani, a young star 30 parsecs from Earth. The planet (artist's impression pictured) is roughly twice the mass of Jupiter and farther away from its star than Saturn is from the Sun, but is comparatively closer than Uranus.

Spectroscopic analysis of light from the planet revealed strong hints of methane in its atmosphere — similar to that found in the giant planets of the Solar System.

Science <http://doi.org/6s6> (2015)



GLACIOLOGY

Asian glaciers make a hasty retreat

The Tien Shan mountain range of Central Asia (pictured) lost more than one-quarter of its glacier mass in the past half-century, and may lose a further 50% by 2050.

Daniel Farinotti of the GFZ German Research Centre for Geosciences in Potsdam and his colleagues used satellite data and glacier models to reconstruct changes in glacier mass and area over the past 50 years, extending previous decadal records. They found that glacier mass

fell by more than 25% from 1961 to 2012 — four times the global average — and glacier area shrank by about 3,000 square kilometres.

As the climate warms, regional summer melt water will probably decrease, putting the water supply at risk in Kyrgyzstan, Kazakhstan, Uzbekistan and parts of China where populations are rising, the authors say.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo2513> (2015)

BIOTECHNOLOGY

Opiates made in yeast from scratch

Engineered strains of laboratory yeast can make small quantities of opiates that are used in pain medications, such as hydrocodone.

Poppy farms are currently the main source of opiates, but production is uncertain because of weather, pests and other factors. Christina Smolke of Stanford University in California and her colleagues introduced more than 20 enzymes into a yeast strain, including some from poppy plants. After 120 hours

of growth using simple sugars, the yeast produced thebaine — which drug makers convert to codeine and other drugs — and hydrocodone.

If the yields could be increased by more than 100,000 times, such microbes could provide an alternative opiate source, the authors say. *Science* <http://doi.org/6s8> (2015)

METABOLISM

Exercise hormone is no myth

A hormone associated with the beneficial effects of exercise has been found in the blood of exercising humans, this time

EVGENY DUBINCHUK/ALAMY

DANIELLE FUTSelaar/Franck Marchis/SETI INST.

thanks to more robust methods than those used in past studies.

Bruce Spiegelman and Steven Gygi of Harvard Medical School in Boston, Massachusetts, and their colleagues reported in 2012 that exercise boosted levels of the hormone irisin in the blood of mice and humans. But other researchers branded the finding a myth, saying that the human gene for irisin cannot produce a protein and that the tests used to detect the hormone were flawed. The team has now repeated the experiment with the more precise technique of tandem mass spectrometry, which confirmed the original findings.

The technique is expensive and slow, but could be used to validate higher-throughput tests for irisin as they are developed, the authors say. *Cell Metab.* <http://doi.org/6s9> (2015)

NEUROSCIENCE

Visual brain area active during sleep

Human brain regions involved in visual awareness are active during certain periods of sleep, as if the people were awake and seeing images.

Yuval Nir at Tel Aviv University, Israel, and his colleagues monitored brain activity in people with epilepsy who had probes implanted as part of their surgical treatment. The researchers gathered data from the brain region involved in visual awareness as the people slept, viewed familiar images on a computer screen or lay awake in a dark room. They found that brain-activity patterns during rapid eye movement (REM) sleep were similar to those that occurred during image viewing — but not to the brain patterns recorded when participants moved their open eyes while lying in darkness.

The eye movements in REM sleep could be directly linked to dreaming rather than just being an indication of a generally aroused brain, the authors say. *Nature Commun.* 6, 7884 (2015)

IMMUNOLOGY

How malaria raises cancer risk

The malaria parasite damages DNA in mouse white blood cells, increasing the chances of them becoming cancerous. This could explain why Burkitt's lymphoma, a cancer of mature B cells, is common in areas where malaria is endemic.

Davide Robbiani and Michel Nussenzweig at the Rockefeller University in New York and their colleagues infected mice with a malaria parasite (*Plasmodium chabaudi*). This caused certain B cells in the spleen to proliferate rapidly and to express high, sustained levels of an enzyme called AID, which breaks DNA.

Malaria alone did not cause cancer in the mice. However, in animals lacking the p53 tumour suppressor gene, malaria promoted a type of lymphoma marked by chromosomal rearrangements similar to those in human Burkitt's lymphoma. *Cell* 162, 727–737 (2015)

CARDIOVASCULAR BIOLOGY

Good cholesterol gets even better

The molecule that removes cholesterol from arteries could have another protective effect in heart disease: curbing inflammation.

High-density lipoprotein (HDL; known as the 'good cholesterol') transports fat from blood-vessel walls to the liver for excretion. But it also carries the fatty signalling molecule S1P, which activates the anti-inflammatory receptor S1P₁. Timothy Hla of Cornell University in New York and his colleagues tested the effect of HDL-bound S1P on cells that line human blood vessels and found that it dampened inflammation. In mice engineered to develop heart disease, those missing the S1P gene had more arterial plaques than did those with the gene.

The researchers think that this further explains the

SOCIAL SELECTION

Popular topics on social media

Trials register sees null results rise

The launch of the clinicaltrials.gov registry in 2000 seems to have had a striking impact on reported trial results, according to a *PLoS ONE* study that many researchers have been talking about online in the past week. A 1997 US law mandated the creation of the registry, requiring researchers from 2000 onwards to record their trial methods and outcome measures before collecting data. The study found that in a sample of 55 large trials testing heart-disease treatments, 57% of those published before 2000 reported positive effects from the treatments. But that figure plunged to just 8% in studies that were conducted after 2000. Study author Veronica Irvin, a health scientist at Oregon State University in Corvallis, says this suggests that registering clinical studies is leading to more rigorous research. Writing on his *NeuroLogica* Blog, neurologist Steven Novella of Yale University in New

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/nlurh8

Haven, Connecticut, called the study "encouraging" but also "a bit frightening" because it casts doubt on previous positive results (go.nature.com/tybzdvd). *PLoS ONE* 10, e0132382 (2015)



ability of HDL to stave off cardiovascular disease. *Science Signal.* 8, ra79 (2015)

CLIMATE CHANGE

Beef farming brings on warming

The most climate-friendly methods of beef production may not be enough to reduce the environmental effects of raising cattle.

Raymond Pierrehumbert at the University of Oxford, UK, and Gidon Eshel at Bard College in Annandale-on-Hudson in New York modelled the warming effects of five different ways of producing beef, including intensive feedlot systems (pictured) and

pasture-based methods. Their analysis was based on the known amounts of greenhouse gases produced by each method.

They found that certain forms of pastured beef production have a smaller climate footprint than feedlot systems. However, because the global population and its appetite for beef will only grow, the footprint of these methods is still sizeable enough that even the most efficient systems will probably not help climate warming to stay below 2°C. *Environ. Res. Lett.* 10, 085002 (2015)

➔ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Access complaint

US federal agencies, including the Food and Drug Administration and the Centers for Disease Control and Prevention, are continuing to restrict journalists' access to information, according to the Society of Professional Journalists in Indianapolis, Indiana. In July 2014, the society complained that staff at the agencies were being prohibited from communicating with journalists and that interviews with federal employees were being monitored. A survey by the society in April this year found that science journalists are also affected by these practices. In a letter sent to US President Barack Obama on 10 August, 53 journalism and open-government groups say that restrictions on access to federal information are dangerous and should be lifted.

FUNDING

Greek bailout boost

The Greek government will soon allow researchers to apply for research grants for the first time in several years, after the country's third bailout in five years was approved on 14 August. The bailout — which must still be ratified by eurozone parliaments — contains two portions of money earmarked for research. These include €59 million (US\$66 million) for proposals previously frozen in the crisis and €53 million for new calls and international projects. The debt crisis has left researchers reliant on international grants — some have had to ask collaborators abroad for basic supplies. It is hoped that the funding will help to address Greece's brain drain. See page 273 for more.



YUE YUEWEI/XINHUA PRESS/CORBIS

Chinese chemicals blast hits port city

China's prime minister, Li Keqiang, has promised a thorough investigation into explosions that occurred on 12 August at a chemicals warehouse in the port city of Tianjin. The blasts killed more than 100 people, with dozens still missing; most of the dead and missing are firefighters. Local officials confirmed that highly toxic sodium cyanide was present at the site, although they said that it

would be neutralized. The blasts also damaged the building that houses China's supercomputer, Tianhe-1, which has been used extensively in the country's space programme. The computer was running smoothly after the explosions, but was shut down for security reasons. Local media aired allegations that regulatory impropriety allowed the warehouse to be sited too close to residential and business areas.

EVENTS

Muslim climate call

Islamic leaders have told the world's 1.6 billion Muslims that it is their religious duty to meet the challenge of climate change. The declaration, made on 18 August at the International Islamic Climate Change Conference in Istanbul, Turkey, called for climate education in mosques and Islamic schools, and appealed to rich, oil-producing nations to lead emissions cuts. The announcement was accompanied by messages of solidarity from other faiths. Some observers are cautiously optimistic about its impact on the behaviour of

individuals, but more sceptical about its ability to influence government policy.

Record El Niño

An El Niño event developing in the eastern Pacific Ocean could be one of the strongest on record, forecasters with the US National Oceanographic and Atmospheric Administration announced on 13 August. The periodic weather pattern — signalled by warming waters and weakening trade winds — often leads to wetter winters in the southern United States and dry conditions in Australia, Indonesia and India. In California, where drought has decimated state water supplies, the rains would be welcome,

but forecasters say that one strong El Niño will probably not slake the state's thirst. See go.nature.com/tjgdhz for more.

Fossil-fuel plea

Almost 1,000 health professionals and academics have signed a letter asking two of the world's biggest medical charities to sell off investments in fossil-fuel companies. The letter, released on 14 August, calls on the Bill & Melinda Gates Foundation in Seattle, Washington, and the London-based Wellcome Trust to divest their shares in the world's 200 largest fossil-fuel companies within 5 years. According to 2014 financial

LEE THOMAS/ALAMY documents, the Wellcome's £18-billion (US\$28-billion) endowment included investments totalling more than £500 million in several such companies. Its director, Jeremy Farrar, has previously defended these investments, arguing that they give the charity leverage to influence the companies' practices.

RESEARCH

Gamer danger

Exposure to violent video games is linked to aggressive behaviour in older children, adolescents and young adults, according to a report issued by the American Psychological Association (APA) on 13 August. The organization reached its conclusion on the basis of 31 studies published in 2009–13. It could not, however, determine the relationship between video games and criminal behaviour or delinquency. In response to the findings, the APA has adopted a resolution that calls for a refined rating system for video games to better reflect violent content.

Ice delays science

The Canadian Coast Guard's *Amundsen* icebreaker (pictured), which is being used for the Canadian part of an Arctic oceanographic programme, is due to end the first leg of a disrupted cruise



this week. The research trip is part of the international GEOTRACES programme studying pan-Arctic Ocean geochemistry. Soon after the *Amundsen* left Quebec City on 10 July, the Coast Guard diverted it to break ice for commercial vessels in Hudson Bay, despite objections from the cruise's chief scientist, Roger François of the University of British Columbia in Vancouver. Although it lost two weeks, the cruise was able to salvage nearly all of its science, but François says that the decision underscores "systemic inadequacies" in Canadian oceanographic research.

Pollution deaths

Air pollution in China is responsible for nearly 17% of the country's deaths each year, according to a study by Berkeley Earth in California, a climate-science

non-profit organization (see go.nature.com/svbild). The report analysed 4 months of data from more than 1,500 pollution-monitoring sites around China. It found that, on average, 38% of the population breathes air deemed unhealthy by US Environmental Protection Agency standards. The estimate that air pollution kills 1.6 million people in China annually is an increase of 400,000 over previous estimates.

BUSINESS

Insect investment

Synthetic-biology firm Intrexon announced on 10 August that it will buy Oxitec, a company in Abingdon, UK, that is developing genetically modified insects. The deal is worth US\$160 million. Oxitec creates sterile *Aedes aegypti* mosquitoes, which are

COMING UP

23–27 AUGUST

Ecologists gather in Manchester, UK, to discuss resilient ecosystems at the 6th World Conference on Ecological Restoration. www.ser2015.org

23–28 AUGUST

Eger, Hungary, hosts the world's microscopists as they gather for the 12th Multinational Congress on Microscopy. www.mcm2015.com

24–28 AUGUST

The 10th Annual Immunotherapy and Vaccine Summit is held in Boston, Massachusetts. www.imvacs.com

released in Brazil to reduce the size of the mosquito population and slow the spread of dengue. Intrexon, in Germantown, Maryland, has also invested heavily in AquaBounty, a biotechnology company in Maynard, Massachusetts, that makes genetically modified salmon.

Genome venture

Genome editing continues to attract heavy interest from investors. Editas Medicine, which was founded in 2013 to develop therapies using the gene-editing technique known as CRISPR, announced on 10 August that it had raised US\$120 million. Investors in the Cambridge, Massachusetts, company include inventor and philanthropist Bill Gates and Google Ventures, the venture-capital arm of the technology giant in Mountain View, California. The versatility of CRISPR has fuelled optimism about its potential applications in agriculture and medicine.

► NATURE.COM

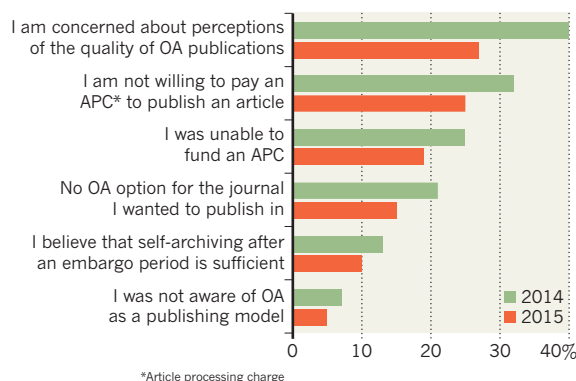
For daily news updates see: www.nature.com/news

TREND WATCH

Scientists' concerns about open-access (OA) publishing might be fading, according to the results of a large Nature Publishing Group survey released on 13 August (see go.nature.com/t2cxrq). The poll found that more than one-third of scientists have not published any OA papers in the past three years, a proportion that has shifted little since 2014. But last year, 40% of those researchers expressed concerns about perceptions of the quality of OA publications; this year, that proportion dropped to 27%.

OPEN-ACCESS PERCEPTIONS CHANGING

Survey results reveal how scientists' attitudes towards open-access (OA) publishing have shifted in the past 12 months.



NEWS IN FOCUS

SOCIETY Gay scientists feel more accepted than non-scientist peers **p.275**

CLIMATE CHANGE Overestimate of China's emissions presents research challenge **p.276**

OLYMPICS Conservationists protest proposed site of Beijing ski runs **p.278**



HEALTH Legalization adds urgency to questions around marijuana **p.280**

AYHAN MEHMET/ANADOLU AGENCY/GETTY



Greek Prime Minister Alexis Tsipras addresses parliament during the vote on a bailout deal in Athens on 14 August.

ECONOMIC CRISIS

Greek bailout set to free up research funds

Research minister says two chunks of money from the European Union can now be released.

BY ALISON ABBOTT

When the banks closed in Greece at the start of July, Vassilis Gorgoulis, a cancer researcher at the University of Athens, was poised to start experiments that journal referees had requested for a high-profile paper. To his horror, he was unable to buy the items he needed because money could no longer be sent out of the country. "I was lucky," he says. "My collaborators at the University of Manchester ordered them for me and sent me the kits I needed."

Greek scientists who do not have collaborators abroad are finding it even harder to carry on doing experiments. The constraints on the flow of capital come on top of massive budget and salary cuts in the five years since Greece's debt crisis began. And they are accompanied

by a stultifying increase in bureaucracy because all public spending in the country has come under intense scrutiny (see *Nature* **517**, 127–128; 2015).

On 14 August, the Greek Parliament and the European Council of Ministers both approved the third bailout of Greece's current economic crisis. That, along with slight improvements in the country's financial liquidity, clears the way for the release of two large chunks of research money, says Costas Fotakis, the Greek research minister. One of them will go towards calls for research proposals to be put out in October — the first major calls for several years.

"This is very encouraging indeed," says Gorgoulis. Without these calls, researchers have had to rely on money from international grants. Moreover, Fotakis, a prominent laser physicist who was appointed to his post in

January when the left-wing Syriza party took power (see *Nature* <http://doi.org/6s4>; 2015), says that he has plans to reinvigorate research, including creating Greece's first fund dedicated to research and innovation.

However, unhappiness with the deal within the Syriza party may lead to snap elections and introduce more uncertainties.

Both chunks of cash come from European Union (EU) 'structural-fund' packages — subsidies for poor regions that can be used for research. One chunk, amounting to €59 million (US\$66 million), had not been released from the structural-fund round that ended in 2014, says Fotakis. It will provide funding to back calls for proposals that had been frozen because of the crisis. The other, for €53 million, is part of a package that will be rolled out over the next 7 years. As well as funding new calls, the ▶

► second chunk of money will cover fees to international research organizations and pay for HEAL-Link — a government-paid portal that provides scientists with access to electronic journals. The portal was closed down on 1 July owing to a lack of funds (see *Nature* <http://doi.org/6s5>; 2015).

Details of the upcoming calls have not been released, but Fotakis says that they will focus on proposals that are designed to address immediate priorities, such as slowing or reversing a crippling brain drain and making the research environment more attractive. But even more important, he says, are his plans for the “day after the crisis”. “We are designing policies for a new scientific landscape ready for when things start to normalize here,” he says.

One planned task is to present Parliament with a law designed by his ministry to make research more efficient by removing bureaucratic obstacles that emerged as a result of austerity. Those include checks on spending insisted on by the organizations that represent Greece’s creditors — even ordering stationery is now a complicated procedure. “This law is ready to go and will be put before Parliament just as soon as the agreement between Greece and its creditors is certain,” Fotakis says.

Panayiota Poirazi, a computational biologist at the Institute for Molecular Biology and Biotechnology in Heraklion, knows only too well that researchers face greater problems than money. She has a €1.5-million ‘starting grant’ from the European Research Council but has not always been able to spend it because of the capital controls. “When the banks closed we could not replace an air-conditioning unit in our computer server room,” she says.

The research and innovation fund that Fotakis plans to create would be fed by public and private sources and foster basic and applied research, Fotakis says. Science will help to promote a “knowledge-based economy”, he says, “but also curiosity-driven research is needed to make an impact on the economy in the long term”.

It might be difficult to find much private investment in Greece right now, he acknowledges, but he is negotiating for support from various sources. One of those is the European Fund for Strategic Investment, which was created earlier this year to stimulate investment in high-risk, high-return innovations (see *Nature* <http://doi.org/6s7>; 2015).

Gorgoulis likes the sound of Fotakis’s plans: “This is the way to go if we want to have a long-term perspective, if Greece is to have a future.” But some researchers worry that the new uncertainties hanging over the current government could end up delaying, or derailing, the research plans. ■

Additional reporting by Anastasia Christakou.



The number of nations with a top-level infectious-disease lab is set to grow as Japan joins the fold.

INFECTIOUS DISEASE

Japan to upgrade biolab security

Fears about Ebola helped forge deal to elevate existing facility to BSL-4 status.

BY HELEN SHEN

Japan is set to join an elite club with its decision to upgrade an existing infectious-disease lab to handle the most-hazardous pathogens. The move sweeps away more than three decades of political opposition to operating a top-biosafety-level facility 30 kilometres west of Tokyo in the city of Musashi-Murayama.

An agreement reached on 3 August between Japan’s health ministry and the mayor of Musashi-Murayama clears the way for the facility to begin limited work with pathogens such as the Lassa and Ebola viruses. Japan’s National Institute of Infectious Diseases (NIID) built the biosafety-level-4 (BSL-4) lab in 1981, but it has been limited to operating as a BSL-3 lab because of community safety concerns. Fears that Ebola might reach Japan during last year’s outbreak in West Africa partly motivated the policy change.

The deal sets several conditions for the lab’s activities: the NIID has committed to maintain

transparency in reporting lab operations and any accident, and the lab must restrict its BSL-4 work to diagnosing and treating patients instead of running a broader research programme. However, virologist Ayato Takada at Hokkaido University in Sapporo, Japan, hopes that the agreement will ease the way for other facilities where scientists can perform basic infectious-disease research at the BSL-4 level. Discussions are under way to build a bigger and more modern BSL-4 lab at Nagasaki University — a move that has similarly met with community opposition.

For now, Takada’s studies of Ebola virus must be done in collaboration with international partners; he flies to the United States several times a year to perform BSL-4 experiments. “It’s time-consuming and expensive,” he says. “I really hope the decision at Murayama will have a good influence on the likelihood of the Nagasaki University BSL-4 plan.”

Upgrading operations from BSL-3 to full BSL-4 status may take several months, says virologist Masato Tashiro, former director of the NIID’s Influenza Virus Research Center.

The original BSL-4 infrastructure at Musashi-Murayama has been maintained over the decades, but new protocols will need to be established and staff trained for the

CHRISTIAN CHARISUS/DPA/PRESS ASSOCIATION IMAGES

higher-security-level work. The lab will also need to import samples from other countries to build a reference library for diagnosing suspected infections.

Typical BSL-3 labs include two sets of self-locking doors and directional airflow to prevent the escape of potentially lethal, airborne pathogens. But BSL-4 labs have extra features that protect the workers and prevent the escape of highly lethal microbes that cause infections for which there are no treatments or vaccines.

EBOLA RESPONSE

Over the years, the NIID made various attempts to gain public support for BSL-4 operations at the site, but some citizens and politicians were concerned that the risks outweighed the benefits. That began to change last year as the Ebola outbreak raged in West Africa. Japan, like many other countries, worried that the virus could cross its borders through international travel.

Japan joins a growing club of nations that have considered or expanded BSL-4 capabilities over the past decade — many spurred by the 11 September 2001 terrorist attacks. The Federation of American Scientists in Washington DC, a non-profit organization specializing in security matters, estimated in 2011 that roughly 40 BSL-4 labs existed or were under construction worldwide, although some researchers consider that an overestimate.

“It was odd that a global player in science and a highly developed industrialized country such as Japan has not had that type of facility. It’s brilliant that they have one now,” says virologist Paul Duprex at Boston University School of Medicine in Massachusetts.

Others, however, argue that BSL-4 labs serve only a small slice of infectious-disease research, and that existing facilities already exceed global needs. “BSL-4 facilities are fantastically expensive to construct and operate, and that comes at the cost of other areas of biomedical research,” says molecular biologist and biosecurity specialist Richard Ebright at Rutgers University in Piscataway, New Jersey. Although the Musashi-Murayama lab itself is limited in scope, Ebright says, “the most likely impact is that this will influence more nations to join this club.” ■

SOCIETY

Most gay scientists are out in the lab

Female-dominated fields found to be more welcoming.

BY MARLA BROADFOOT

Lesbian, gay, bisexual and transgender (LGBT) scientists feel more accepted in the workplace than their peers in other professions, a US survey suggests. The study, published in the *Journal of Homosexuality*, also found that respondents in scientific fields with a high proportion of women, such as the social sciences, were more likely to be out to their colleagues than those in male-intensive disciplines such as engineering (J. B. Yoder and A. Mattheis J. *Homosexuality* <http://doi.org/6vc>; 2015).

That result may suggest that laboratories with more women tend to be more receptive to people who do not fit the stereotype of a scientist as a straight, white man, says Jeremy Yoder, an evolutionary ecologist at the University of British Columbia in Vancouver, Canada, and co-author of the study.

The Queer in STEM survey gauged the experiences of more than 1,400 LGBT individuals working in science, technology, engineering and mathematics (STEM) fields. The 58-item online survey included questions about gender and sexual identity, professional expertise and whether respondents felt that their work and social communities were welcoming or hostile.

Other studies have indicated that out employees feel more accepted at work, whereas those who are closeted or not open about their identities are more likely to be stressed at work and have negative attitudes towards their work and co-workers (K. H. Griffith and M. R. Hebl J. *Appl. Psychol.* **87**, 1191–1199; 2002).

Participants rated their openness about their gender and sexual identities on a scale from 0 (“I am not out to anyone in this group”) to 5 (“As far as I’m aware, everyone in this group

could know”). The results indicated that the majority of the respondents (57%) were out to half or more of their colleagues, a greater proportion than the 47% of people who said that they were out in a 2014 survey of the general US workforce by the Human Rights Campaign Foundation in Washington DC.

“There is reason to believe this difference could be real,” says Yoder. “In STEM workplaces you are working with a fairly well-educated set

“You are working with a fairly well-educated set of co-workers, and you may very well be able to expect a more open culture.”

of co-workers, and you may very well be able to expect a more open culture.”

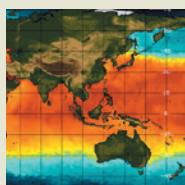
Most respondents identified themselves as lesbian, gay, bisexual, transgender, queer or asexual. Some ticked more than one box, and

5% wrote down an identity not included in the survey’s list, such as polyamorous, non-monogamous or pansexual. The researchers relied on social media and networking to recruit respondents, so they were concerned about the representativeness of the sample. A geographic analysis found that respondents came from US Census Bureau regions roughly in proportion to their estimated LGBT populations, but survey respondents tended to be young and to work in the life sciences.

“The study breaks new ground in providing a national portrait,” says Kristen Renn, a social scientist at Michigan State University in East Lansing who studies LGBT college students. She adds that the findings “provide a sense of possibility” to LGBT and queer youth who are exploring careers. “For those wondering, ‘Can I be an openly queer chemist?’” she says, the answer is: “Yes. And you won’t be alone.” ■



TOP STORY



Developing El Niño weather system could be among the strongest on record go.nature.com/tjgdhz

MORE NEWS

- Frogs can ‘switch on’ genetic resistance to pesticides go.nature.com/ume21e
- Sickness increases diversity in fruit-fly offspring go.nature.com/wzewvd
- US judges block attempt to stop retractions go.nature.com/v9ymae

NATURE PODCAST



Lessons from Hurricane Katrina; China’s emissions lower than thought; and borrowed genes nature.com/nature/podcast



China's cement industry may produce less carbon than was thought.

CLIMATE CHANGE

China's emissions overestimated

Analysis could force climate scientists to revisit estimates of global carbon cycle.

BY JEFF TOLLEFSON

China's carbon emissions may be significantly lower than previously thought — about 14% less in 2013 than estimated by the Chinese government and others, according to research published this week in *Nature* (Z. Liu et al. *Nature* **524**, 335–338; 2015). The analysis draws on data from more than 4,200 Chinese mines — including new measurements of the energy content of coal — among other sources.

“At the beginning of the project we thought that the emissions might be higher” than existing estimates, says Zhu Liu, an ecologist at Harvard University in Cambridge, Massachusetts, and lead author of the study. “We were very surprised.”

His team's findings do not unseat China from its position as the world's largest emitter of carbon dioxide. Even when the lower estimate is taken into account, China's carbon output for 2013 is still more than two-thirds higher than that of the United States, the second-largest emitter. But the study underscores longstanding uncertainties in the methods with

which scientists calculate the emissions of individual nations, and how much carbon cycles through the atmosphere and into oceans and ecosystems. For comparison, the cumulative reduction in Chinese emissions outlined in the study — roughly 2.9 billion tonnes from 2000 to 2013 — is larger than the estimated amount of carbon that the world's forests pulled out of the atmosphere from 1990 to 2007.

That presents a problem for researchers who study the carbon cycle. “We can easily go back and retroactively adjust Chinese carbon-emission estimates,” says Ashley Ballantyne, a climate scientist at the University of Montana in Missoula. “Unfortunately, we cannot go back and adjust all the previous studies on the global carbon cycle and their conclusions based on the previously biased emission estimates.”

The Chinese government releases data on energy consumption and production at the provincial and national levels, but those statistics often conflict with each other, and are revised frequently. Liu and his team analysed government data on energy production and on exports and imports of coal, oil and gas. They found that China's fossil-fuel use was

10% above the official government estimate, but that the country's overall emissions were lower once China's reliance on low-quality coal from domestic mines was taken into account. This is because lower-quality coal contains less carbon than higher-quality deposits, so burning it produces less energy and less heat-trapping CO₂.

The team says that its estimate for how much CO₂ will be produced by burning Chinese coal is around 40% less per unit than the figures adopted by the Intergovernmental Panel on Climate Change. And the team calculates that emissions from cement production, a coal-fuelled process that is a major contributor to global emissions, are 45% below existing estimates.

The coal measurements were collected from mine reports and from a project sponsored by the Chinese Academy of Sciences that assesses the country's cumulative carbon emissions and carbon uptake by ecosystems across China. Liu says that the quality of Chinese coal is likely to be getting worse as the country burns through its best reserves.

“This is probably the best available estimate of emissions from coal burning in China, and that is an important contribution,” says Gregg Marland, a geologist at Appalachian State University in Boone, North Carolina, and a co-author of the study. But he adds that the revised figure is within the range of uncertainty reported in existing inventories.

And it may need to be increased as the Chinese government releases further energy data, says Glen Peters, a climate-policy researcher at the Center for International Climate and Environmental Research in Oslo. Although China said in February that its coal consumption had dropped between 2013 and 2014, he notes, the government has since increased its cumulative estimate of coal consumption over the past decade by 12–14%. Scientists are still waiting for the government to release revised

“With Chinese energy statistics, there is always a ‘but’.”

estimates of energy production, including imports and exports, over the past decade. Liu says that his team's estimates are unlikely to change when the latest data are released later

this year, but Peters says that the figures may need to rise by as much as 7%.

Such uncertainty arises in part from the many different ways to define and measure energy consumption; researchers do not know what kind of assumptions the Chinese government is making with its data. “If China reported their CO₂ emissions, then we would know the assumptions that they want to make and many of these issues would then go away,” says Peters. The challenge, he adds, is unlikely to be resolved anytime soon.

“With Chinese energy statistics,” he says, “there is always a ‘but’.” ■

XIEZHENG/CORBIS

MATERIALS SCIENCE

Superconductor high spurs wave of physics

Researchers strive to reproduce and understand zero-resistance material that works at a record high temperature.

BY EDWIN CARTLIDGE

Hydrogen sulfide — the compound responsible for the smell of rotten eggs — conducts electricity without any resistance at a record high temperature of 203 kelvin (-70°C), reports a paper published on 17 August in *Nature*¹.

The work is a step towards finding a long-sought room-temperature superconductor. The first results were released on the arXiv preprint server in December²; more followed in June³. Both have already sparked a wave of follow-up calculations and experiments, not to mention excitement.

The discovery is “historic”, says physicist Fan Zhang of the University of Texas at Dallas, adding that its impact will be “far ranging”. A superconductor that works at room temperature could improve electricity generation and transmission, as well as giving a boost to current uses of superconductivity, such as in the huge magnets used in medical imaging machines. In a News & Views article⁴ that accompanies the *Nature* paper online, Igor Mazin of the Naval Research Laboratory in Washington DC describes the finding as “the holy grail of superconductors”.

The results are the work of Mikhail Erements, Alexander Drozdov and their colleagues at the Max Planck Institute for Chemistry in Mainz, Germany. They find that when they subject samples of hydrogen sulfide to extremely high pressures — around 1.5 million atmospheres (150 gigapascals) — and cool them below 203 K, the samples display the classic hallmarks of superconductivity: zero electrical resistance and a phenomenon known as the Meissner effect, in which a material is placed in an external magnetic field but there is no field inside the sample, unlike in normal materials.

DEFYING CONVENTION

Other scientists are intensely interested in the result because it was achieved without using exotic materials such as copper-containing compounds called cuprates, says physicist Christoph Heil of the Graz University of Technology in Austria. Until now, these have held the record for the highest superconducting temperature — 133 K (-140°C) at ambient pressure and 164 K (-109°C) at high pressure.

Many scientists had assumed that superconductivity at temperatures of more than a few tens of kelvins required such materials. But Heil says that the pressurized hydrogen sulfide seems to be a ‘conventional’ superconductor — vibrations in the material’s crystal lattice drive electrons to form ‘Cooper pairs’ that can flow through the crystal without resistance. And in calculations reported in April⁵, Matteo Calandra of the Pierre and Marie Curie University in Paris and his colleagues found that the results could indeed be explained using a modified version of the conventional theory of low-temperature superconductivity based on lattice vibrations.

For others, such theoretical analyses are superfluous until the result by Erements and co-workers is confirmed experimentally by independent teams. Several are working towards that goal, including Katsuya Shimizu of Osaka University in Japan and his colleagues, who have seen the loss of resistance in pressurized hydrogen sulfide, but have yet to observe the Meissner effect. Meanwhile, three Chinese groups and one US team contacted by *Nature* say that they have yet to confirm either the electrical or magnetic effects.

If Erements and colleagues are proved right, then other compounds that pair hydrogen with, for instance, platinum, potassium, selenium or tellurium instead of sulfur may also be good candidates for high-temperature superconductivity, according to several other research groups. Taking a different tack, Zhang in Dallas and Yugui Yao of the Beijing Institute of Technology in China predict⁶ that substituting 7.5% of the sulfur atoms in hydrogen sulfide with phosphorus and upping the pressure to 2.5 million atmospheres could raise the superconducting transition temperature all the way to 280 K, which is above the freezing point of water. ■

1. Drozdov, A. P., Erements, M. I., Troyan, I. A., Ksenofontov, V. & Shylin, S. I. *Nature* <http://dx.doi.org/10.1038/nature14964> (2015).
2. Drozdov, A. P., Erements, M. I. & Troyan, I. A. Preprint at <http://arxiv.org/abs/1412.0460> (2014).
3. Drozdov, A. P., Erements, M. I., Troyan, I. A., Ksenofontov, V. & Shylin, S. I. Preprint at <http://arxiv.org/abs/1506.08190> (2015).
4. Mazin, I. *Nature* <http://dx.doi.org/10.1038/nature15203> (2015).
5. Errea, I. *et al. Phys. Rev. Lett.* **114**, 157004 (2015).
6. Ge, Y., Zhang, F. & Yao, Y. Preprint at <http://arxiv.org/abs/1507.08525> (2015).



A model of the alpine skiing area displayed during Beijing's bid to host the 2022 Olympic Winter Games.

CONSERVATION

Chinese biologists lead Olympics outcry

Proposed alpine skiing area lies within nature reserve.

BY DAVID CYRANOSKI

Celebrations of Beijing's success in its bid to host the 2022 Olympic Winter Games have been tempered by protests from young biologists over the proposed alpine ski site. They say it falls within a protected national nature reserve that contains many rare species, including Beijing's only Shanxi orchids (*Cypripedium shanxiense*).

More importantly, say other biologists, the proposed site would violate environmental protection laws touted by the government, and create a precedent that would hamper already fraught efforts to conserve other, more significant sites.

Local officials have since said that they will redraw the lines that delineate the reserve — but this has left the protesters unsatisfied. Meanwhile, their online posts on the subject, to the Chinese social media website Weibo, are no

longer visible; the protesters say they have been blocked.

On 31 July, the International Olympic Committee, based in Lausanne, Switzerland, announced that Beijing would host the 2022 Winter Olympics, beating its only rival, Almaty in Kazakhstan, by 44 votes to 40. But nature enthusiasts quickly noted that the proposed alpine skiing routes fall within the core area of the 4,600-hectare Songshan National Nature Reserve. On 1 August, Wang Xi, who recently received his PhD in plant biology and works at the Chinese Academy of Sciences in Shanghai, overlaid maps from the International Olympic Committee's evaluation report with those from the reserve's website and posted the result on his Weibo account: both the start and end of the alpine runs fall within the reserve, he found (see 'Olympic overlap').

Xi told *Nature's* news team that his main motivation was to spread news of the possible

ecological impact on plants there, including on three orchid species that are classified at the highest protection level under Beijing's conservation system. "It's a chance for the government to connect with the people and talk to each other to solve this problem," he says. "I am not against the Olympic Games, but they should be carried out in an environmentally friendly way."

WEIBO CLICKS

Xi's post was clicked some 240,000 times and forwarded more than 1,000 times, but within two days he found that it no longer appeared online. Word spread anyway. Yun Ji, a 28-year-old entomologist in Beijing who did not want his affiliation mentioned, had copied Xi's maps. He added information on and pictures of the various plants found in the reserve that are on a list of the rarest and most rigorously protected of Beijing's plants, along with photos of birds whose habitat, he says, could also be destroyed. His Weibo post is now inaccessible from the site — he claims it disappeared within four hours — but while it was online he says that it was forwarded 3,000 times, and one of those reposts was forwarded an additional 16,000 times.

Neither the Chinese Olympic Committee nor the Beijing government has released official statements responding to the allegations in the posts; in the International Olympic Committee evaluation report, the skiing area is described as being "adjacent to" the reserve. But on 7 August, the deputy mayor of Yanqing, the county in Beijing that has jurisdiction over the Songshan National Nature Reserve, announced that the borders of the reserve would be "adjusted" to take in land to the northeast of the current boundaries, and that some of the original reserve will be taken "to provide the necessary space for local sustainable development and to promote interaction between ecological protection and economic society".

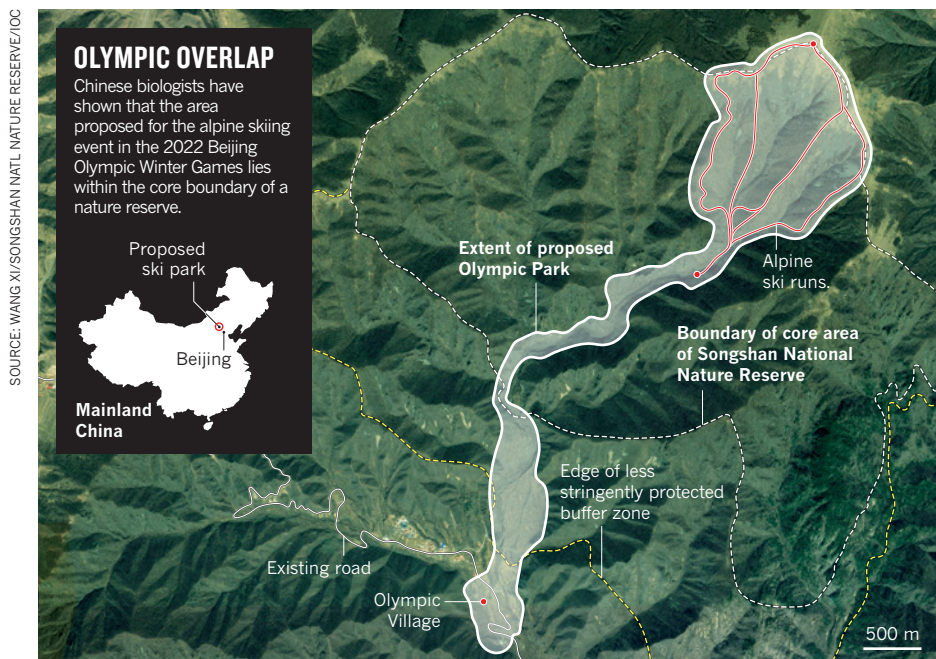
The details are not yet clear, but the proposed reserve is 31% larger than the original and does not overlap with the ski slopes, according to interviews with the deputy mayor published

"I am not against the Olympic Games, but they should be carried out in an environmentally friendly way."

by Chinese media outlets. The mayor's office deferred queries from *Nature's* news team to the bid committee, but the team did not receive a response from that committee nor from

the China Olympic Committee. The International Olympic Committee has also not replied to queries from *Nature's* news team on how it handles applications that have inconsistencies, and whether conflict with a nature reserve would have affected its decision.

According to 2013 Chinese government regulations, those who wish to change



environmental sustainability and conservation seems to be backing away from its promises. In May, for example, China's environment ministry released a notice, signed by ten government agencies, that stated that any development at odds with a reserve's function is "strictly forbidden". Gu and Xi both worry that failure of the strict policies in Beijing would send a broader signal to local governments. If Beijing violates Songshan's reserve, "it will be easier for local governments to give construction projects higher priority than conservation issues", says Gu, whose Weibo posts on the subject are no longer accessible. "The real impact", he says, would be "the breaking of Chinese laws and policies on nature reserves".

Xi says that a better site for the alpine skiing event would be Zhangjiakou, a city in neighbouring Hebei province, which will also host some Winter Olympic events. Zhangjiakou has mountains that are already developed and would suffer no environmental loss, he says. ■

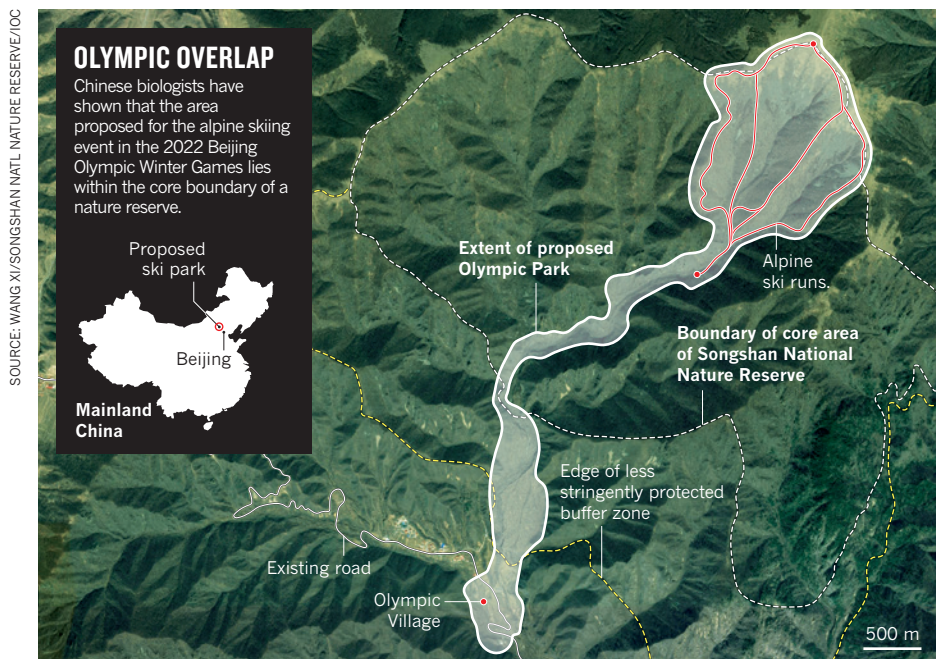
CLARIFICATION

The News story 'Stanene makes its debut' (*Nature* **524**, 18; 2015) should have made it clear that the stanene was made by experimentalists at Shanghai Jiao Tong University.

nature-reserve boundaries must submit an application that includes public comment, an ecological assessment and four other documents. Xi and Ji say that they can find no evidence that this happened.

Lei Gu, a postdoctoral researcher in evolution

and conservation biology at Peking University in Beijing, says that Songshan is not that important in terms of biodiversity. The bigger problem, he says, is that a government that has been increasingly issuing statements and regulations that emphasize its commitment to



environmental sustainability and conservation seems to be backing away from its promises. In May, for example, China's environment ministry released a notice, signed by ten government agencies, that stated that any development at odds with a reserve's function is "strictly forbidden". Gu and Xi both worry that failure of the strict policies in Beijing would send a broader signal to local governments. If Beijing violates Songshan's reserve, "it will be easier for local governments to give construction projects higher priority than conservation issues", says Gu, whose Weibo posts on the subject are no longer accessible. "The real impact", he says, would be "the breaking of Chinese laws and policies on nature reserves".

Xi says that a better site for the alpine skiing event would be Zhangjiakou, a city in neighbouring Hebei province, which will also host some Winter Olympic events. Zhangjiakou has mountains that are already developed and would suffer no environmental loss, he says. ■

CLARIFICATION

The News story 'Stanene makes its debut' (*Nature* **524**, 18; 2015) should have made it clear that the stanene was made by experimentalists at Shanghai Jiao Tong University.

nature-reserve boundaries must submit an application that includes public comment, an ecological assessment and four other documents. Xi and Ji say that they can find no evidence that this happened.

Lei Gu, a postdoctoral researcher in evolution

and conservation biology at Peking University in Beijing, says that Songshan is not that important in terms of biodiversity. The bigger problem, he says, is that a government that has been increasingly issuing statements and regulations that emphasize its commitment to



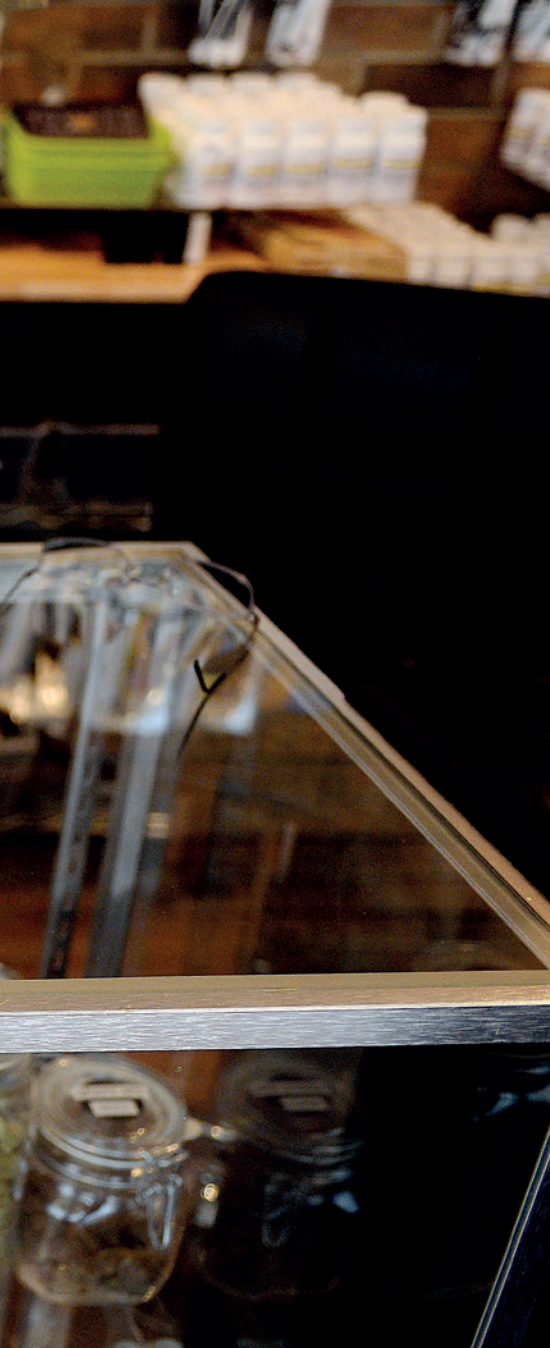
THE CANNABIS EXPERIMENT

As marijuana use becomes more acceptable, researchers are scrambling to answer key questions about the drug.

BY DANIEL CRESSEY

In 2013, Beau Kilmer took on a pretty audacious head count. Citizens in the state of Washington had just voted to legalize marijuana for recreational use, and the state's liquor control board, which would regulate the nascent industry, was anxious to understand how many people were using the drug — and importantly, how much they were consuming.

The task was never going to be straightforward. Users of an illicit substance, particularly heavy users, often under-report the amounts they take. So Kilmer, co-director of the RAND Drug Policy Research Center in Santa Monica, California, led a team to develop a web-based survey that would ask people how often they had used cannabis in the past month and year. To help them gauge the amounts, the surveys included scaled pictures showing different quantities of weed. The survey, along with other data the team had collected, revealed a rift between



Retailers in Colorado can now sell cannabis to the public.

have gone further, legalizing the drug for recreational consumption. A handful of other states including California and Massachusetts are expected to vote on similar recreational-use measures by the end of 2016.

But the rapid shift has caught researchers on the back foot. “Broadly speaking, there’s about 100 times as many studies on tobacco or alcohol as there are on illegal substances,” says Christian Hopfer, a psychiatry researcher at the University of Colorado School of Medicine in Denver. “I don’t think it’s the priority it should be.”

Despite claims that range from its being a treatment for seizures to a cause of schizophrenia, the evidence for marijuana’s effects on health and behaviour is limited and at times conflicting. Researchers struggle to answer even the most basic questions about cannabis use, its risks, its benefits and the effect that legalization will have.

The quick shifts in policies should provide a plethora of natural experiments, but the window will not be open for long. “There’s an opportunity here. Some of the most informative research we can do is right at the moment the market changes,” says Robert MacCoun, a social psychologist and public-policy researcher at Stanford Law School in California who worked with Kilmer on the research done in Washington.

WHAT ARE THE NEGATIVE EFFECTS?

For years, the debate over the drug’s safety has been polarized. Those seeking legalization claim that it is basically harmless. Yet governments around the world have placed cannabis among the most-dangerous illegal drugs, running hard-hitting campaigns warning of the threats that it poses to mental health and social well-being.

Scientists are fairly sure about some things, particularly when it comes to the short-term effects. They know, for instance, that it impairs memory and coordination, and can cause paranoia and psychosis². These are some of the classic symptoms of being ‘high’ and can have major health effects in and of themselves. Studies have found, for example, that drivers are between twice and seven times as likely to crash if they have recently smoked the drug^{3,4}.

In the longer term, effects are less clear, but there are a few that most scientists agree on. Contrary to many popular arguments, there is evidence that cannabis is addictive. Around 9% of users become dependent on the drug, showing signs of addiction such as developing tolerance or experiencing withdrawal symptoms when they stop using. Beyond that, however, long-term effects have been difficult to pin down.

Cannabis is often smoked, and this can raise the risk of respiratory problems and

possibly lung cancer. A 2008 study in New Zealand found that smoking pot increased the risk of lung cancer by 8% for each ‘joint-year’ (the equivalent of smoking a joint per day for one year), even after taking tobacco use into account⁵. But other studies have found little to no correlation with lung cancer, even for heavy users⁶.

Other health outcomes are even more difficult to disentangle from confounding factors. Some researchers have found links to poor educational performance, low social attainment — such as job status — and altered brain development. For example, the Christchurch Health and Development Study — which followed almost 1,300 children born in New Zealand in 1977 — found that people who used cannabis daily are around 50% more likely to have psychotic symptoms⁷ than are non-users and are at greater risk of not finishing school⁸.

And another study from New Zealand, which followed 1,000 people in Dunedin from birth to age 38, shows that persistent cannabis use, especially if started young, correlates with steeper declines in IQ in later

“Marijuana research is like tobacco research in the ’60s. Any study about harms is challenged.”

life and with problems with memory and reasoning compared with people who have never used the drug⁹.

Hall says that the association with negative social and mental-health outcomes has been consistently observed, but the debate “is how we explain that association”, which he says will probably involve a combination of factors.

The difficulty, says Valerie Curran, a psychopharmacologist at University College London, lies in teasing apart correlation and causation, because “there are so many confounders”. For example, adolescents who use cannabis are probably also drinking excessive amounts of alcohol and engaging in other risky activities. Attributing the effects to one particular substance or behaviour is therefore very difficult.

Similar problems abound in the hotly contested link between cannabis and schizophrenia. Multiple studies have shown an increased risk of this mental-health disorder

perception and reality. Based on prior data, state officials had estimated use at about 85 tonnes per year; Kilmer’s research suggested that it was actually double that, about 175 tonnes¹. The take-home message, says Kilmer, was “we’re going to have to start collecting more data”.

Scientists around the world would echo that statement. Laws designed to legalize cannabis or lessen the penalties associated with it are taking effect around the world. They are sweeping the sale of the drug out of stairwells and shady alleys and into modern shopfronts under full view of the authorities. In 2013, Uruguay became the first nation to legalize marijuana trade. And several countries in Europe — Spain and Italy among them — have moved away from tough penalties for use and possession. Thirty-nine US states plus Washington DC have at least some provisions for medicinal use of the drug. Washington, Colorado, Alaska and Oregon

in people who use cannabis versus people who do not. A study of 50,000 Swedish men aged 18–20 found that heavy users were around three times more likely to develop schizophrenia than those who had never used the drug¹⁰. Although the increase in risk was significant, the overall risk is still low — just 1.4% of men who reported using cannabis developed the disorder, compared with 0.6% of those who said they had never tried the drug. Some cannabis advocates suggest that the link may be down to people with such problems ‘self-medicating’, but this is difficult to prove.

Many of the negative health outcomes seem to be exacerbated if the drug is used in adolescence, leading to suggestions that cannabis is adversely affecting developing brains. And effects may also be linked to the drug’s potency, which in itself is hard to pin down.

As cannabis use becomes legal, the data may become easier to collect. But the drug’s use is still low compared with alcohol and tobacco, says Wayne Hall, an addiction researcher at the University of Queensland in Brisbane, Australia, so it is hard to draw firm conclusions. Marijuana may be the most popular illegal drug, he says — about 44% of US adults have used it at some point in their lives according to one source — but only about one in ten have used it in the past year. By contrast, around 70% drank alcohol in that time. “The number of people who use it with any regularity for a long time is pretty small. The longer-term consequences are really understudied,” says Hall.

HOW STRONG IS IT?

A major question for researchers — and a complication in interpreting the evidence — is dosing. There are more than 85 cannabinoid chemicals in pot. The one of most interest to researchers — and users — is tetrahydrocannabinol (THC). Growers have been able to breed high concentrations of the chemical into strains of the plant meant for recreational and medicinal use. A potency-monitoring programme run by the University of Mississippi for the US National Institute on Drug Abuse (NIDA) found that THC levels have steadily increased in the United States¹¹, from 2–3% in 1985–95 to 4.9% in 2010. The increase is even starker for imported cannabis seized by law-enforcement officials. For these drugs, potency has gone from less than 4% in the late 1980s and early 1990s to more than 12% in 2013.

But it is hard to determine the amounts of THC being consumed by the average customer. It is unclear, for example, whether users ‘titrate’ their doses, adjusting their intake according to the potency. Nicotine users are known to do this with cigarettes, but nicotine does not impair judgement in the same way that cannabis does. And the

Reefer madness

After more than a century of fairly lenient laws surrounding cannabis, public attitudes in the United States and elsewhere shifted against the drug in the early 1900s. Anti-marijuana propaganda escalated in the 1930s, with films depicting the horrors that befell users (see right).

1976

The Netherlands decriminalizes possession, use and sale of small quantities of marijuana.

1996

California becomes the first state to legalize medical cannabis, through Proposition 215.

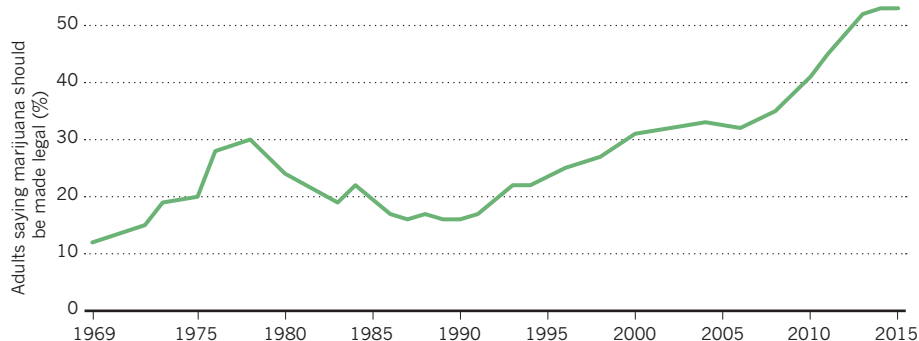
2012

Washington and Colorado vote to legalize recreational marijuana for adults aged 21 or older, although the drug does not become available for sale until 2015.



IMAGE: MOVIE POSTER IMAGE ART/GETTY; DATA SOURCE: PEW RESEARCH CENTER

Survey data have revealed a dramatic shift in attitudes towards marijuana in the United States.



effects of THC are less immediate, especially for edible forms.

The escalating potency raises questions about previous research because users in older studies may have been consuming lower-potency cannabis, and the effects may be different (see ‘Research gaps’). A study published earlier this year, for example, linked high-potency cannabis to a threefold-increased risk of psychosis versus non-use but found no association with lower-potency forms¹². And many researchers have complained that the pot approved for study in experiments funded by NIDA is a poor match for what is used recreationally or medicinally.

In tandem with changing laws, the Colorado Department of Public Health and Environment (CDPHE) is establishing reference labs to check the potency of what is sold. And the US government is expanding the varieties of marijuana that researchers with federal funding can obtain.

In places where the drug is legal, existing labelling standards may also be inadequate. A survey done between August and October last year found that only 17% of edible cannabis products in San Francisco, Los Angeles and

Seattle had accurate labels. More than half had less THC than claimed, and some contained significantly more¹³. “A lot of people get a rude surprise,” says MacCoun.

ARE THERE MEDICAL BENEFITS?

Although states are starting to ease restrictions on recreational use of marijuana, what got the ball rolling in changing public perceptions and the legal landscape for pot were the arguments for its medical use.

Colorado introduced its rules allowing medical marijuana more than a decade before it allowed recreational use. The amendment to the state’s constitution listed eight conditions for which marijuana was approved: cancer, glaucoma, HIV/AIDS, cachexia (a progressive wasting syndrome), persistent muscle spasms, seizures, severe nausea and severe pain. But, says Larry Wolk, executive director and chief medical officer of the CDPHE, “those are dictated by the constitution and not necessarily by medical research”.

Although there is a huge amount of anecdotal evidence — and well-organized advocacy groups that campaign for easier access to medical marijuana — there is little conclusive scientific evidence for many of the

claimed medical benefits. One of the reasons for this dearth of evidence is that money generally has been obtainable only for research on the negative effects of cannabis. That is beginning to change.

When Colorado first legalized the drug, its public-health department began collecting fees from patients who applied to purchase pot at medical dispensaries. By 2014, the state had amassed more than US\$9 million, most of which was ploughed back into a medical marijuana research programme selected by the CDPHE. Among the projects funded by the Colorado millions, there are two investigating whether cannabinoids can help to mitigate seizures in childhood epilepsy. Similar research is being pursued in the United Kingdom and elsewhere in the United States.

Another, more-established use is for people with multiple sclerosis. A cannabis-based spray has been approved in 27 countries for treatment of muscle problems associated with the disease, such as spasms.

Other claimed benefits of marijuana, such as boosting appetite in people with AIDS, are supported by more-limited evidence. If positive effects can be clearly demonstrated, it would be a huge vindication for marijuana advocates. It might also go some way towards justifying medical-marijuana legislation.

In the meantime, however, scientists are watching the emerging cannabis frontier with wary eyes. "I think it's an experiment," says Robert Booth, a psychiatry researcher at the University of Colorado. "When this study is all said and done, we'll know a whole lot about the effects of marijuana."

WHAT HAPPENS WHEN YOU MAKE IT LEGAL?

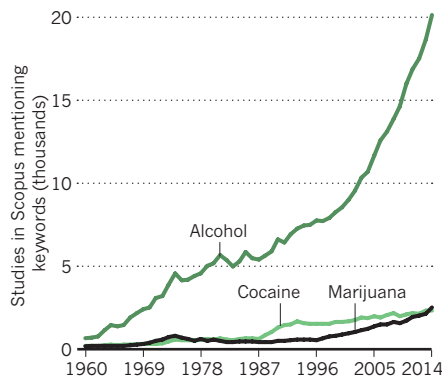
One of the biggest questions is how legalization will change usage patterns. One place in which researchers are looking for answers is Europe, where cannabis regulation tends to be much lighter than it is in the United States (see 'Reefer madness'). In the United Kingdom, some police forces overlook cannabis use and small-scale growing operations. Spain allows private consumption, but still has restrictions on sales.

The most extreme and long-standing example is the Netherlands, which decriminalized the possession and sale of small quantities of cannabis in 1976. But although some streets of Amsterdam have been transformed into pungent tourism hotspots, the country as a whole has not changed its habits much.

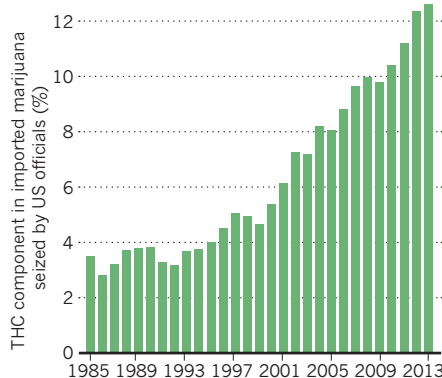
Although hard data on cannabis use in Europe is patchy, the Netherlands does not have hugely more users than other nations. Data aggregated by the United Nations Office on Drugs and Crime put use in the Netherlands at about 7%. That is more than in Germany (5%) and Norway (5%), about the same as in the United Kingdom and less than in the United States (15%). Nor has the Netherlands

Research gaps

Published research on marijuana lags behind that of other drugs, such as alcohol and cocaine, according to a search of the bibliographic database Scopus.



The amount of tetrahydrocannabinol (THC) in cannabis has risen sharply in the past three decades, raising concerns about the relevance of older research and studies using less-potent varieties of the drug.



seen a huge spike in use of harder drugs, dampening fears that marijuana serves as a gateway to more-dangerous substances such as heroin and cocaine. The message from the Netherlands, says Franz Trautmann, a drugs-policy researcher at the Trimbos Institute in Utrecht, the Netherlands, is that "a very liberal policy doesn't lead to a skyrocketing prevalence". Rather, cannabis is endemic, he says. "We can't control this through prohibition. This is something which more and more is recognized."

But the lesson from the Netherlands may be limited because the drug is still illegal, and growing and selling large quantities is still punishable by law. Colorado has gone further by legalizing not merely the drug's use, but the whole production chain, and that could have fundamentally different effects on the economics of pot. "Legalized production really raises the prospect of a dramatic drop in price," says MacCoun. "It's conceivable marijuana prices could drop 75–80% in a fully legalized model." (Although Uruguay legalized the drug in 2013, it reportedly has struggled to regulate production and to set up working dispensaries.)

The effects of a sharp drop in cost are

unknown. Taxation may also have unintended consequences. If states tax by weight, users might look to higher-potency strains to save money. And once cannabis is a business, it gains a business lobby. Cannabis researchers already talk of being bombarded with e-mails from pro-cannabis groups if they make negative comments about the drug. "Marijuana research is like tobacco research in the '60s," says Hopper. "Any study about harms is challenged. It's really something." Many fear that the big money now to be found in cannabis will drive attempts to obfuscate the risks. "If the commercial interests are too big, then the profit interest is prevailing above the health interest. This is what I'm afraid of," says Trautmann.

Legalization provides an opportunity to answer some important questions. In a few years, Colorado, Washington and others will know (if only roughly) how legalization affects usage patterns, the number of car crashes and the number of people seeking help for drug dependency. The CDPHE-funded programmes will have added to the knowledge of beneficial effects. And continuing long-term studies of large groups of users will provide more evidence for statisticians who are attempting to disentangle correlation and causation on the negative impacts.

"When a jurisdiction changes its marijuana laws, that provides an opportunity for greater leverage on the questions of cause and effect," says MacCoun. But, he adds, the signals will only really be clear if the laws result in a dramatic increase in use — something that is neither a given, nor necessarily desirable. "Obviously, we don't want marijuana use to rise just to allow us to answer our questions, but if it does, we'll be poring over all the data." ■

Daniel Cressey writes for Nature from London.

1. Kilmer, B. et al. *Before the Grand Opening: Measuring Washington State's Marijuana Market in the Last Year Before Legalized Commercial Sales* (RAND Corp., 2013); available at go.nature.com/ibu8vl
2. Volkow, N. D., Baler, R. D., Compton, W. M. & Weiss, S. R. B. *N. Engl. J. Med.* **370**, 2219–2227 (2014).
3. Hall, W. *Addiction* **110**, 19–35 (2014).
4. Ramaekers, J. G., Berghaus, G., van Laar, M. & Drummer, O. H. *Drug Alcohol Depend.* **73**, 109–119 (2004).
5. Aldington, S. et al. *Eur. Respir. J.* **31**, 280–286 (2008).
6. Hashibe, M. et al. *Cancer Epidemiol. Biomarkers Prev.* **15**, 1829–1834 (2006).
7. Fergusson, D. M., Horwood, L. J. & Ridder, E. M. *Addiction* **100**, 354–366 (2005).
8. Fergusson, D. M., Horwood, L. J. & Beutrais, A. L. *Addiction* **98**, 1681–1692 (2003).
9. Meier, M. H. et al. *Proc. Natl Acad. Sci. USA* **109**, E2657–E2664 (2012).
10. Zammit, S., Allebeck, P., Andreasson, S., Lundberg, I. & Lewis, G. *Br. Med. J.* **325**, 1199 (2002).
11. Office of National Drug Control Policy *National Drug Control Strategy: Data Supplement 2014* (White House, 2014); available at go.nature.com/mm8qyk
12. Di Forti, M. et al. *Lancet Psychiatry* **2**, 233–238 (2015).
13. Vandrey, R. et al. *J. Am. Med. Assoc.* **313**, 2491–2493 (2015).

COMMENT

NEURODIVERSITY Scientific hubris abounds in sobering history of autism spectrum **p.288**

GENOMICS New play puts Native American dispute centre stage **p.289**

MENTAL HEALTH How Italy revolutionized its psychiatric care **p.290**



CONSERVATION Reclamation destroys coral reefs in South China Sea **p.291**

VINCENT LAFORET/NYT/REDUX/EYEVINE



New Orleans flooded and burning, the week after Hurricane Katrina.

Hurricane Katrina's lessons for the world

Ten years after the devastation of the US Gulf coast, **Edward B. Barbier** calls for coastal protection plans like that adopted by Louisiana for the world's most vulnerable nations.

Ten years ago, Hurricane Katrina devastated the US Gulf coast states. It caused around US\$110 billion in damages, more than 1,800 deaths, and displaced 1.2 million people¹.

The disaster led to a rethink of the management of the Gulf coastline. In the seven decades preceding 2005, Louisiana had lost coastal lands, mainly marshes, totalling around 4,900 square kilometres — an area the size of Trinidad and Tobago².

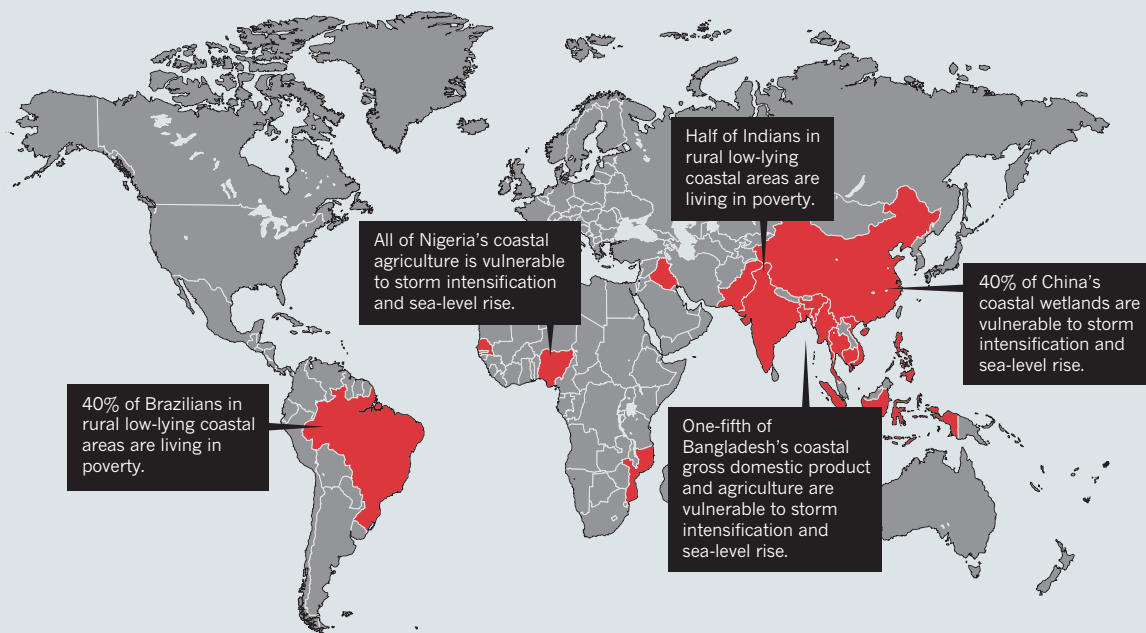
Following the hurricane, the President's Gulf Coast Ecosystem Restoration Task Force recommended extensive wetland re-establishment, noting that they “provide a natural flood attenuation function”³.

After the hurricane season in 2005, the Louisiana state legislature created the Coastal Protection Restoration Authority (CPRA) and tasked it with coordinating the local, state and federal efforts. The CPRA embarked on five-year coastal

master plans to guide policymakers in developing a more sustainable coast. I served on the science and engineering board that oversaw the scientific analysis for the 2012 master plan. This plan was charged with evaluating, prioritizing and integrating coastal protection and restoration projects over the next 50 years, with a total budget of \$50 billion. The plan accounts for how coastal restoration and protection in Louisiana may have to ►

15 MOST-AT-RISK NATIONS

Almost 60 million people live in poverty in the low-lying rural areas of these countries' coastlines.



▶ adapt to differing climate-change and sea-level-rise scenarios.

The 2012 plan is not perfect. Oyster, shrimp and crab harvesters, landowners, waterfowl hunters and shipping interests have expressed concerns about river-diversion projects, delaying them. And the marsh-creation plans are ambitious: around 100 square kilometres of new wetlands from 2012 to 2031 (ref. 2). Combined with the building of levees and other structures, however, these efforts lower the damage costs associated with the risk of flooding along Louisiana's coast by \$5.3 billion to \$18 billion annually².

Flaws and merits aside, Louisiana's 50-year coastal-planning strategy represents a new way of thinking about the long-term management of coasts. Resilience is the new aim — against short-lived natural disasters that have immediate and often extreme impacts, such as flooding and storm surges, and against long-term climatic changes that have more gradual impacts, such as sea-level rise, saline intrusion and erosion.

Other parts of the world urgently need such a long-term strategy. Around 38% of the global population — 2.5 billion people — live within 100 kilometres of the coast. More than three-quarters of these people are in developing countries. The frequency of flooding associated with cyclones and other major coastal storms is likely to increase by the 2050s as a result of accelerated sea-level rise⁴. Coastal areas are the front lines of climate change.

The most vulnerable are poor, rural populations in developing countries that live

less than 10 metres above sea level, in low-elevation coastal zones (LECZs). In 2010, around 267 million people lived in the rural areas of LECZs. By 2100, the figure is projected to be 459 million (ref. 5).

Almost all of the world's 60 million poor people living in rural LECZs reside in just 15 countries: India, Bangladesh, Myanmar, Cambodia, Nigeria, Pakistan, Iraq, Mozambique, Senegal, Brazil, China, Indonesia, the Philippines, Vietnam and Thailand (see '15 most-at-risk nations'). Many of these people depend on agriculture and fishing, and rely on natural barriers such as mangroves for protection against storms and other coastal hazards that are set to intensify with climate change. Yet, as happened on the Gulf coast, these barriers are being eroded (see Supplementary Information; go.nature.com/3mxhjlj).

These 15 nations should be the priority for a long-term global planning strategy to protect coasts and populations that are vulnerable to damaging storms.

THREE LINES

Consider Bangladesh. The current capacity, infrastructure and coping strategies of its coastal communities are inadequate to deal with current storms, let alone any of increased intensity. Over the next 35 years, the area of the coastline that is vulnerable to flooding of deeper than 3 metres could increase by more than 69%; the zones vulnerable to floods that are deeper than 1 metre could grow by 14%⁶.

A one-off investment of an estimated \$2.4 billion is needed to cope with these hazards in Bangladesh from now until 2050, with extra maintenance and operating costs of \$50 million per year. These sums are half the projected financial costs of future damages, deaths and injuries from the increased hazards associated with climate change⁶.

A long-term global resilience strategy for coasts needs three lines of defence. First is the restoration of 'green' infrastructure — robust coastal ecosystems, such as salt marshes, oyster and coral reefs, mangroves, seagrass beds and beaches. These are essential, especially in remote, rural parts of developing countries^{7,8}. Second is the right 'grey' infrastructure, such as seawalls, dikes, barrages and diversions as well as raising the elevation of buildings and flood-proofing structures. Third is to involve key local stakeholders throughout; this builds institutional and coastal community capability.

This third element — engagement, participation and empowerment — must not be overlooked. A good example of engagement is India's Joint Mangrove Management programme that, with the help of non-governmental organizations, enlists local communities in conservation and afforestation efforts⁷. Similarly, Louisiana's Coastal Master Plan was developed in consultation with state legislators, coastal parish officials, community groups, business leaders, advocacy organizations and Louisiana scientists. A series of public meetings attracted more than 750 people to comment on the plan;



Indian villagers participate in mangrove restoration.

more than 2,200 comments came in online, by e-mail and by post².

In the long term, community development will also need to include policies and investments that drive livelihood diversification — moving away from coastal resource-based activities such as fishing and agriculture towards manufacturing and services. And these communities need investments and improvements in health, education, local governance and institutions. As a first step towards determining which long-term investments should be a priority in Kenya, Madagascar, Mauritius, the Seychelles and Tanzania, for example, assessments have been conducted of coastal communities' vulnerability and adaptability to the main impacts of climate change on coral-reef fisheries⁹.

Finally, the most vulnerable populations may need to be encouraged to migrate to non-coastal areas. Schemes such as voluntary property acquisition or assisted relocation may be required, as was recommended by the Louisiana master plan². In Bangladesh and other developing countries, improving roads in remote, rural LECZs might be the best option to encourage voluntary migration from areas of high flood risk to safer areas with better employment opportunities⁶.

"The most vulnerable populations may need to be encouraged to migrate to non-coastal areas."

ADAPTATION FUNDS

The 2012 Louisiana master plan has demonstrated that it is possible to put all three lines of coastal defence in place — green, grey and social. But it has done so in one of the richest nations in the world. Developing countries will clearly need international support to create and implement such plans. For instance, the Global Environment Facility and the United Nations Development Programme (UNDP) are already collaborating with the government of Bangladesh in support of its five-year, \$8.5-million Community-based Adaptation to Climate Change through Coastal Afforestation programme⁸.

Such long-term planning efforts and capacity-building should be financed through international climate-change adaptation funds. This could be done on a cost-sharing basis with relevant governments and local jurisdictions in coastal zones. Currently, around \$13.4 billion in the UN's Green Climate Fund as well as the main adaptation funds have been pledged to assist developing countries adapt to climate change. Just one quarter of that could make rural LECZs more resilient, and improve the livelihoods of nearly 60 million poor people currently living in these zones.

The next step should be for the COP21 UN climate summit in Paris this December to earmark multilateral adaptation funds to support developing countries in their efforts to devise a long-term global planning strategy for enhancing and protecting

coasts and populations vulnerable to damaging storms. The World Bank, UNDP and the UN Environment Programme should also commit to assisting local and national authorities through a coordinated effort. ■

Edward B. Barbier is professor of economics at the University of Wyoming in Laramie, Wyoming, USA. His book *Nature and Wealth: Overcoming Environmental Scarcity and Inequality* will be published this year.
e-mail: ebarbier@uwyo.edu

1. Knabb, R. D., Rheme J. R. & Brown, D. P. *Tropical Cyclone Report: Hurricane Katrina* (US National Hurricane Center, 2005).
2. Coastal Protection and Restoration Authority of Louisiana. *Louisiana's Comprehensive Master Plan for a Sustainable Coast* (Office of Coastal Protection and Restoration, 2012).
3. Gulf Coast Ecosystem Recovery Task Force. *Gulf of Mexico Regional Ecosystem Restoration Strategy* (Gulf Coast Ecosystem Recovery Task Force, 2011).
4. Woodruff, J. D., Irish, J. L. & Carnargo, S. J. *Nature* **504**, 44–52 (2012).
5. Center for International Earth Science Information Network. *Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2* (NASA Socioeconomic Data and Applications Center, 2013); available at <http://dx.doi.org/10.7927/h4mw2f2j>
6. Dasgupta, S. et al. *Vulnerability of Bangladesh to Cyclones in Changing Climate: Potential Damages and Adaptation Cost* (World Bank, 2010).
7. DasGupta, R. and Shaw, R. 2013. *Ocean Coast. Mgmt* **80**, 107–118 (2013).
8. Rawlani, A. K. & Sovacool, B. K. *Mitig. Adapt. Strat. Glob. Change* **16**, 845–863 (2011).
9. Cinner, J. E. et al. *Glob. Environ. Change* **22**, 12–20 (2012).



Sensory support is helping this man with autism spectrum disorder to connect with his therapist.

AUTISM

Seeing the spectrum entire

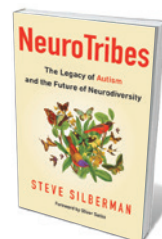
Chris Gunter examines a comprehensive history of the science and culture surrounding autism studies.

Within two generations, the popular and scientific understanding of autism spectrum disorder (ASD) and related conditions have undergone a massive shift in some parts of the world. We have moved from routine institutionalization (or worse) of people with ASD to an appreciation of a spectrum of social communication.

How did this sea change come about? Journalist Steve Silberman has been writing and commenting on autism for years, notably with a 2001 feature in *Wired* magazine on ASD rates in California's Silicon Valley. He has compiled his exhaustive research into *NeuroTribes* to try to answer that question.

Genetic and neurological studies now firmly suggest that 'autism' describes a constellation of behaviours. This is defined in the fifth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5, 2013)

as "persistent difficulties in the social use of verbal and nonverbal communication" and "restricted, repetitive patterns of behavior, interests, or activities". In 2014, the US Centers for Disease Control and Prevention estimated that 1 in 68 US children had ASD. Co-occurring challenges abound — including anxiety and digestive problems — but only about half of people diagnosed with ASD also have intellectual disability. We now recognize that autistic people can have a hugely varying range of impairments and abilities, which can change throughout life. This recognition



NeuroTribes: The Legacy of Autism and the Future of Neurodiversity
STEVE SILBERMAN
Avery: 2015.

has been hard won. Silberman takes us back to the early twentieth century and clinics in Europe, Russia and the United States that first began using the terms autism or autistic psychopathy (from the Greek word *autos*, meaning 'self') to describe apparent withdrawal into an inner world.

The clinicians most often credited with discovery are Hans Asperger and Leo Kanner. Asperger's clinic in 1930s Vienna embraced the full range of ASD. But Silberman asserts that to protect his charges from euthanasia by the Nazis, Asperger focused his case reports on gifted children ostracized by their peers, later termed high-functioning. Eventually, these cases would be called Asperger's syndrome; in the DSM-5, controversially, this diagnosis is folded into ASD.

Kanner wrote a 1943 paper about his patients in Baltimore, Maryland, which is generally considered the first description of autism (*Nerv. Child* 2, 217–250; 1943). He focused on "extreme autistic aloneness" and "an anxiously obsessive desire for the maintenance of sameness". This restricted view would shape the field for five decades.

With child psychologist Bruno Bettelheim, Kanner did further damage in the 1940s and 1950s by popularizing the unfounded idea that loveless "refrigerator" parenting caused autism. The pair suggested that children with autism be institutionalized 'for their own good'. Silberman provides ample biographical detail on many such personalities in early autism research, as well as Bernard Rimland, who founded the US National Society of Autistic Children (now the Autism Society) in 1965 and pioneered the involvement of parents in support networks and the search for treatments and educational strategies.

As he must, Silberman discusses the disastrous impact of the now-retracted 1998 paper in which UK surgeon Andrew Wakefield (since struck off the medical register) alleged a link between ASD and the measles, mumps and rubella vaccine. Silberman demolishes the study's claims, long dismissed as fraudulent and debunked by massive longitudinal studies.

Instead, rising rates of ASD diagnosis in the past four decades (US prevalence was estimated at 1 in 5,000 in 1975) can be tied to — but not wholly explained by — the use of expanded diagnostic criteria from the 1980s. UK clinician Lorna Wing and her colleagues, whose revisions to the DSM-III were adopted in 1987, fully expected to see a rise in diagnoses, hearkening back to Asperger's 1944 definition of "autistic psychopathy", in which he described his observations as "not at all rare". Wing adopted the term autism spectrum, recalling the words of statesman Winston Churchill: "Nature never draws a line without smudging it."

The rise in diagnoses is also linked to awareness increased by films featuring autistic

characters, such as *Rain Man* (1988), as well as the advent of the Internet and social media, which enabled the formation of support groups and allowed communication in ways that can be easier for people with ASD. As some of the stigma has ebbed, parents have become more willing to seek diagnoses for their children and even themselves.

One of the most powerful talks at the 2015 International Meeting for Autism Research was by John Elder Robison, who has been diagnosed with Asperger's syndrome. He told the room of scientists and clinicians,

"As the stigma has ebbed, parents have become more willing to seek diagnoses."

"The reason diagnostic labels are so important to those of us with autism is that without those labels, we only have the labels we got in the streets, which are hateful." While we debate diagnostic criteria, we must remember that there are people receiving these labels (and not) who deserve respect, understanding and support.

Fittingly, *NeuroTribes* ends with the neurodiversity movement that is now emerging. Groups such as the Autism Self Advocacy Network campaign for spending less on finding 'cures' and more on designing appropriate support and accommodation for people with ASD and their families and caregivers.

Any work on ASD will be seen as a triumph by some and a travesty by others; Silberman has opted not to be shy. Great contributions include interviews with prominent scientists and self-advocates. He intersperses these with his own opinion, for example portraying the *DSM-III* as a way for psychiatrists to link to "Big Pharma". His affection for detail can get in the way, as in a chapter on the development of radio and electronic bulletin boards. Readers should look elsewhere for a primer on what we do and do not understand of the basic biology of ASD; that is not Silberman's intent.

NeuroTribes is no casual read. The parent of a newly diagnosed child looking for information may be put off by the weight of past horrors on show. But for people in the field, or anyone seeking to understand the interplay between medical science and patient communities, it should be an essential resource. My own copy is already dog-eared and underlined throughout. ■

Chris Gunter is director of communications for the Marcus Autism Center at Children's Healthcare of Atlanta in Georgia. She is also at Emory University School of Medicine in Atlanta. e-mail: drchrisgunter@gmail.com

GENOMICS

Bioethics on stage

Brendan Maher reviews a play inspired by a famous clash between geneticists and a Native American tribe.

In the early 1990s, researchers at Arizona State University (ASU) in Tempe collected blood samples from the Havasupai people, an isolated Native American tribe living in the Grand Canyon. A decade later, tribe members sued the university for misusing the samples, on the grounds that they had not been fully informed of the study's scope. The legal battle culminated in a US\$700,000 payout from ASU; the remaining samples were returned.

That much is true. But much else about the story is disputed. It has surfaced in articles and books as an example of the cultural sensitivities that come up in research. Some argue that it has become a fable of arrogant scientists riding roughshod over people's rights — a tale that has hardened mistrust between researchers and Native American groups.

The latest retelling is the drama *Informed Consent*, playing at the Duke on 42nd Street in New York City. Seasoned playwright Deborah Lauffer sees her work as an exploration of truth and the clash between science and religion. Although the scientist at its centre comes across as a caricature of hubris, the piece effectively presents some ethical, legal and social complexities of modern genomics.

It is highly fictionalized. The real researchers were anthropologist John Martin and geneticist Therese Markow, who in the 1990s were trying to find genetic underpinnings for the Havasupai's high incidence of type 2 diabetes — roughly 55% among women and 38% among men. In the play, Ken (played by Jesse Perez) is a social anthropologist who has worked with an unnamed tribe in the Grand Canyon for 40 years, and

Informed Consent
DEBORAH ZOE LAUFFER
The Duke on 42nd Street, New York City.
Until 13 September.

Jillian (Tina Benko) is an ambitious genetic anthropologist struggling with the knowledge that she carries a genetic mutation that ensures that she will develop early-onset Alzheimer's disease.

Jillian jumps at Ken's offer to run a study looking for diabetes markers in tribe members' DNA. Soon, she reveals other intentions, such as studying how the tribe came to America — even though this would conflict with their own story about where they came from. Ken emphasizes that this is "strictly" a diabetes study. The fourth wall fading momentarily, Jillian tells the audience that he never said "strictly". Other cast members jump in to support her version, and Ken revises his line — a playful swipe at the indefinite nature of remembered truth.

Jillian convinces tribe members to hand over samples of their blood — which they deem sacred — by suggesting that DNA tests are their only hope of halting the diabetes epidemic. She also deludes herself into thinking that the work will lead to funding to pursue a cure for Alzheimer's. Her ambition is driven by fear, for herself and for her daughter Natalie (DeLanna Studi), who has a 50% chance of inheriting the mutant gene. In a parallel plot line, Jillian and her husband fight over whether to test Natalie.

The play crescendos as a tribal spokesperson (also played by Studi) confronts Jillian at a talk on the tribe's early migrations out of East Asia, saying that she had no right to use the samples to study this. Under threat of a lawsuit, Jillian's university fires her and the remaining samples are ceremonially returned.

In the mid-2000s, when the real case came to wide attention, progress on techniques for studying DNA was outpacing understanding of how the research might affect participants. There was no clear evidence that Markow or Martin had broken any rules, but the case and others prompted reconsideration of informed-consent documents, which are meant to lay out the risks for participants.

Lauffer acknowledges the limits in presenting this story with certainty, especially in an engaging stage play, which *Informed Consent* most definitely is. She notes: "I guess what happened is much less important than what we can learn from the outcomes." ■

Brendan Maher is a features editor at Nature.



DeLanna Studi (left) and Tina Benko in *Informed Consent*.

Italy's psychiatric renaissance

Andrea Tone admires the history of a radical shift in the treatment of mental illness.

Up to the 1960s, people in Italy's asylums were stripped of their possessions and personhood: victimized by rigid rules, overcrowding and filth, tied to beds or caged and separated from the outside world by towering walls. By the late 1970s, the landscape had been transformed by a reform movement led by 'critical psychiatrist' Franco Basaglia. Historian John Foot explores this extraordinary revolution and its legacy in *The Man who Closed the Asylums* — first published in Italian as *La "Repubblica dei Matti"* (*The 'Republic of the Mad'*; Feltrinelli, 2014).

The movement, Foot shows, came to condemn asylums as morally bankrupt and of limited therapeutic benefit. Spearheaded by radical psychiatrists and joined by journalists, intellectuals, patients, activists and politicians, it laid the foundation for a reform act, Law 180. Passed in 1978 by health commissioners rather than Parliament, the law authorized the de-institutionalization of patients, the establishment of community-based psychiatric services and the closure of asylums. That process finished only in the late 1990s, but it returned 100,000 patients to society. For the first time, a broad coalition united in rejection of asylums to parlay its views into a national political mandate.

The revolution began in Gorizia. An isolated city on the border of what was then Communist Yugoslavia, Gorizia was politically centre-right: mostly Christian Democrats, with some neo-fascist groups. It was an unlikely site to spark a revolution.

Its asylum, like others in Italy, was "dark and sinister," a place of "horror," notes Foot. In 1961, Basaglia became its director, primed for change. After spending six months in prison for anti-fascist politics under Benito Mussolini's Fascist government, Basaglia knew how institutional power could thwart individual liberties. His critique of asylum psychiatry was influenced by philosopher Michel Foucault and sociologist Erving Goffman, as well as psychiatrist Maxwell Jones, whose work turning UK asylums into open therapeutic communities gave Basaglia, and his wife Franca Ongaro, a new paradigm.

In Gorizia, Basaglia assembled a team — the "équipe" — of physicians and reformers including Giovanni Jervis and Ongaro, the latter of whom Foot credits as the principal author of Basaglia's myriad publications. By the mid-1960s, the équipe had opened up the asylum. Freed from shackles, patients rose and slept when they pleased. They socialized at the institution's new bar (equipped with a



Franco Basaglia in Colorno, Italy, in 1971.

jukebox), saw films, played sport and gave voice to their experiences in a patient-run magazine, *Il Picchio* (*The Woodpecker*). Many took day trips and mixed with locals. Patients and staff met daily to discuss their care and to debate broader questions. Were reforms the goal, or a prelude to closing the asylum? How should tensions between asylum leaders and local authorities — not all of whom embraced having patients in their midst — be resolved?

Despite ideological divisions within and beyond the asylum, Gorizia became famous. Activists demanding freedom and societal change visited to witness "new forms of democracy in action". *The Negated Institution* (Einaudi, 1968), a best-selling book by the collective with Basaglia as editor, detailed the transformation and propelled critical psychiatry into the national spotlight.

By then, Basaglia and others had come to believe that patients would be better off in communities with decentralized psychiatric services. He left Gorizia along with some members of his team, and the "Gorizia diaspora", along with the popularity of *The Negated Institution*, had a ripple effect. Psychiatric care was restructured in other Italian cities, including Reggio Emilia, Parma



The Man Who Closed the Asylums: Franco Basaglia and the Revolution in Mental Health Care
JOHN FOOT
Verso: 2015.

and especially Trieste. Basaglia's work there culminated in its asylum's closure in 1980 — the year that Basaglia died of a brain tumour. Trieste became a centre of pilgrimage for critical psychiatrists and a model for reform recognized by the World Health Organization.

The Man Who Closed the Asylums is ambitious and well researched. It is also detailed and long, packed with block quotes and a large cast of characters, as well as a smattering of case studies of asylums and local politics, making it an occasionally choppy read. It is less a biography of Basaglia than a contextualization of the man and the movement in a network of historical actors and the politics and ideology of post-war Italy.

Foot says little about how patients were diagnosed, the spectrum of their disorders and medical interventions, particularly after psychiatric drugs such as the anti-psychotic chlorpromazine reached Italy in the 1950s. Scholars such as Joel Braslow and Andrew Scull have generated a vast literature on these questions, which Foot sidesteps. Also missing is a discussion of what non-asylum psychiatric care, if any, existed in Italy before Law 180.

Foot's argument would have benefited from more comparative analysis. Before Gorizia, many North American asylums had introduced reforms. In 1956, for instance, the use of anti-psychotics encouraged US public asylums to discharge more people than they admitted. Yet, as Foot avers, no country or movement went as far as Italy in recalibrating the organization and politics of mental health care. Law 180 (and broader health reforms that soon subsumed it) was unevenly enforced, failed to create a singular model of care and was undercut by political opposition. Yet Foot's impassioned story reminds us that the future is neither immutable nor ordained, and that small groups of people in peripheral places can change history. ■

Andrea Tone is professor of history and Canada Research Chair in the Social History of Medicine at McGill University in Montreal. She is the author of *The Age of Anxiety*. e-mail: andrea.tone@mcgill.ca

Correspondence

Halt reef destruction in South China Sea

China's land reclamation in the South China Sea has raised geopolitical concerns (see go.nature.com/tmrpih). It is also destroying coral reefs on a large scale — together with their fragile ecosystems.

This loss of reef habitat is estimated from high-resolution satellite imagery to be about 1,400 hectares so far in the Spratly Islands (<http://amti.csis.org/island-tracker>). The reefs are among the most biologically diverse on the planet. They are home to threatened turtle and bird species, a crucial stopover for migratory birds, and a source of larvae for the South China Sea's overexploited coastal fisheries. Further dredging will kill nearby sediment-dwelling communities and sea-grass nursery habitats that take up carbon.

The global scientific, conservation and legal communities must unite to halt this destruction. Reef loss can be publicly monitored with Google Earth imagery and tools, enabling international governing bodies to enforce China's obligations under international law to protect and sustainably manage their reefs (see go.nature.com/wdzf9k). And funding should be fast-tracked for the proposed protective Spratly Islands Marine Peace Park (J. W. McManus *et al.* *Ocean Dev. Int. Law* **41**, 270–280; 2010).

Elizabeth M. P. Madin
Macquarie University, Sydney, Australia.
dr.elizabeth.madin@gmail.com

STEM teaching: the need for wider skills

The innovations to science, technology, engineering and mathematics (STEM) education you recommend (*Nature* **523**, 272–274 and 282–284; 2015) have already had a major impact on the teaching of medicine in Europe over the past 15 years, especially in the United

Kingdom and the Netherlands. These have been inspired by initiatives such as Best Evidence Medical Education (BEME; www.bemecollaboration.org) and ASPIRE (www.aspire-to-excellence.org), both run by the Association for Medical Education in Europe.

There have been 32 systematic reviews examining the literature to determine which teaching methods work and why, and how they could be made more effective — including three highly cited BEME reviews (S. A. Azer *Acad. Med.* **90**, 1147–1161; 2015). The ASPIRE initiative has ensured that teaching is now recognized internationally alongside excellence in research (see R. M. Harden and D. Wilkinson *Med. Teach.* **33**, 95–96; 2011).

There has also been a drastic change in performance assessment of medical students. This no longer relies on regurgitating facts. Examples include the objective structured clinical examination, workplace-based assessments and entrustable professional activities, which emphasize the importance of feedback in monitoring and supporting undergraduate learning.

Trudie E. Roberts *University of Leeds, UK; and Association for Medical Education in Europe, Dundee, UK.*

Ronald M. Harden *Association for Medical Education in Europe, Dundee, UK.*
t.e.roberts@leeds.ac.uk

STEM teaching: the need for wider skills

Now more than ever, students of science, technology, engineering and mathematics (STEM) need to develop skills for real-world problem solving (see *Nature* **523**, 272–274; 2015). These include leadership, collaboration, communication and interdisciplinary thinking — all sought after by companies and organizations that must compete

in today's global economy.

In the Integrated Concentration in Science (iCons) Program I have launched at the University of Massachusetts Amherst (see www.cns.umass.edu/icons-program), STEM students explore the part that science can play in tackling societal challenges such as cancer, climate change and clean energy. For example, one iCons team identified electrical energy storage as a crucial societal need, which motivated them to design and perform an original investigation on the efficient electrolysis of water. Such innovative educational experiences generate new categories of learning that broaden the meaning of student success — important both for students and for the future.

Scott Auerbach *University of Massachusetts Amherst, USA.*
auerbach@chem.umass.edu

Funders must not coerce scientists

The reported complicity of the American Psychological Association (APA) in the 'enhanced interrogation' programme of the US Department of Defense (DOD) represents improper exploitation by the DOD of its relationship with the APA in pursuit of its own objectives, distorting scientific ethics and practices in the process (see *Nature* **523**, 255; 2015). The scientific community as a whole could learn from the mechanisms that the DOD used to influence the APA, as reported in an independent review (go.nature.com/4vpdob).

Historically, similarly inappropriate relationships have sometimes developed between granting agencies and scientists (see, for example, S. Reverby *Examining Tuskegee* (Univ. North Carolina Press, 2009) and *Nature* **467**, 645; 2010). Scientists' alignment with institutional preferences is often a force for good, but vigilance

is needed to prevent erosion of ethical or evidential standards.

Institutional coercion such as that revealed in the report is disastrous for the public perception and integrity of science. The moral force of research standards is weakened if they are subject to political concerns.

We all need to address this potential for abuse without impeding the normal functioning of scientists and their granting institutions. For example, negotiations on ethics and research practices between institutions could be overseen to ensure good-faith participation and eliminate any back-door dealing. Such regulatory authority could be absorbed by existing bodies such as the US Federal Inspectors General.

Cole Helsell *University of California, San Francisco, USA.*
cole.helsell@ucsf.edu

India's inspiring former president

A. P. J. Abdul Kalam (1931–2015) was the first great scientist to become president of India (see *Nature India* go.nature.com/x8rltb). He died on 27 July while delivering a lecture on 'The liveable planet' to students at the Indian Institute of Management in Shillong.

Dubbed the 'missile man' because of his pioneering years as head of India's missile programme, he was a big supporter of fundamental and applied science in the country. He was a favourite among students, igniting the minds of primary-school children and PhD students alike about science through his passion and approachability, patience and simplicity.

Bhaswati Chatterjee *National Institute of Pharmaceutical Education and Research, Hyderabad, India.*

Suman S. Thakur *Centre for Cellular & Molecular Biology, Hyderabad, India.*
sst@cceb.res.in

CELL BIOLOGY

Tagged tags engage disposal

Damaged organelles called mitochondria were once thought to be disposed of by a simple signalling cascade. Cell-based analyses now reveal that a network of complicated molecular interactions initiates disposal. [SEE ARTICLE P.309](#)

NORIYUKI MATSUDA & KEIJI TANAKA

Mitochondria are the energy-producing factories of the cell, so it is essential for cellular health that they function properly. When mitochondria are damaged, for example by stress or ageing, the electrical potential across their membranes decreases. These depolarized mitochondria are then marked with the molecule ubiquitin, which tags them for disposal. Disruption of this process might lead to the accumulation of damaged mitochondria, and has been implicated in Parkinson's disease¹. The molecular connection between ubiquitin and mitochondrial disposal has been controversial, but in a paper on page 309, Lazarou *et al.*² demonstrate that this interaction is mediated by phosphorylation.

Unwanted or damaged components of the cell are digested or recycled en masse by an intracellular degradation process known as autophagy. Increasing evidence indicates that selective forms of autophagy (such as mitophagy, which targets only mitochondria) enable intracellular quality control by specifically disposing of one type of damaged component. During mitophagy, the kinase enzyme PINK1 and the ubiquitin ligase enzyme Parkin cooperate to attach ubiquitin tags to proteins on the outer membrane of depolarized mitochondria¹ — a process called ubiquitylation.

In 2010, it was proposed³ that a linear signalling cascade leads to the clearance of damaged mitochondria. In this model, PINK1 accumulates on mitochondria in response to a decrease in membrane potential, and recruits Parkin, which catalyses the ubiquitylation of outer mitochondrial membrane (OMM) proteins. Autophagy receptor proteins such as p62 then assist in degradation by binding to both the ubiquitylated substrate and to components of the autophagy machinery, such as the protein LC3. LC3 coats structures called phagophores, which engulf mitochondria in preparation for autophagic degradation.

However, this simple model has since been questioned, for several reasons. First, although the disappearance of OMM proteins was initially used to monitor mitophagy, later papers^{4,5} reported that OMM proteins ubiquitylated by Parkin are subject to degradation through a different pathway. Second,

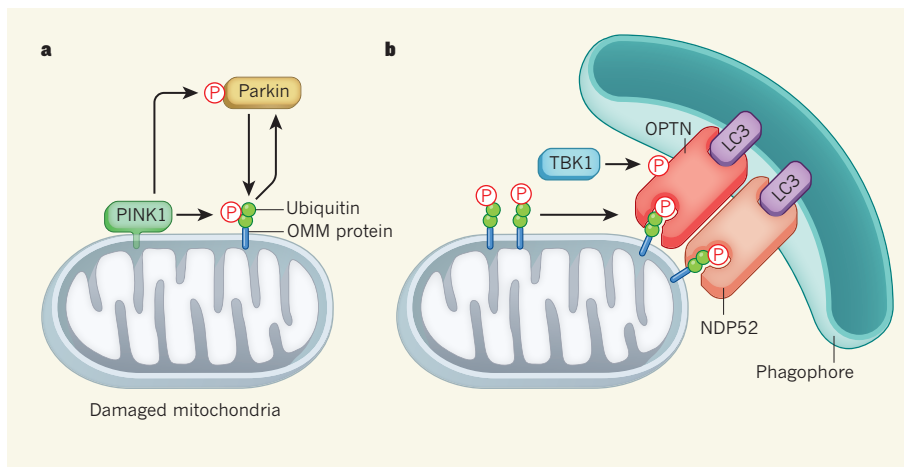


Figure 1 | The molecular mechanism of mitophagy. **a**, During mitophagy, the kinase enzyme PINK1 accumulates on damaged mitochondria, phosphorylating (P) the ubiquitin molecules that tag outer mitochondrial membrane (OMM) proteins. PINK1 also phosphorylates the enzyme Parkin, and PINK1 and ubiquitin both activate and recruit Parkin to mitochondria. Parkin adds more ubiquitin to damaged mitochondria, thereby amplifying the mitophagy signal. **b**, Lazarou *et al.*² add to this already complex model by demonstrating that phosphorylated ubiquitin recruits the proteins OPTN and NDP52. These autophagy receptors bind to phosphorylated ubiquitin and to proteins such as LC3 that coat structures called phagophores — in the case of OPTN, this depends on it being phosphorylated by the enzyme TBK1. Phagophores then engulf the mitochondria in preparation for disposal.

the drug normally used to induce depolarization-dependent mitophagy in experimental conditions, CCCP, was found to cause other cellular changes that also increase the amount of autophagy⁶. Third, the contribution of p62 to PINK1–Parkin-mediated mitophagy has proved controversial^{7,8}.

Over the past two years, several papers^{9–15} have revealed that PINK1 phosphorylates ubiquitin and that phosphorylated ubiquitin (phospho-ubiquitin) is both a receptor for and an activator of Parkin (Fig. 1a). These results contradict the simple hierarchy of the old model, and instead suggest that the components of the pathway can react interdependently. Ubiquitin might act upstream of Parkin as well as downstream, for example, because it can activate Parkin when phosphorylated by PINK1. And PINK1 might act downstream of Parkin as well as upstream because Parkin-catalysed ubiquitin tags can be substrates for PINK1 phosphorylation.

To further investigate the interactions between components of the mitophagy cascade, Lazarou *et al.* studied mitophagy in cells that

lacked genes encoding five autophagy receptors (p62, NBR1, NDP52, TAX1BP1 and OPTN), either individually or in combination. To ensure that mitophagy was monitored accurately, the authors used degradation of mitochondrial DNA, rather than degradation of OMM proteins, as an indicator of mitophagy. As a second indicator, they analysed spectral shifts in mitochondrial fluorescent protein (shifts occur when mitochondria are engulfed for degradation). To depolarize mitochondria, the researchers used the drugs oligomycin and antimycin A, which target mitochondria more specifically than do pan-depolarizing agents such as CCCP.

Lazarou and colleagues found that p62 and NBR1 are not essential for PINK1–Parkin-mediated mitophagy, but that NDP52 and OPTN (and to some extent, TAXBP1) are required for mitophagy, a finding that is partly consistent with previous studies^{7,8,16}. The authors also showed that the ubiquitin-binding domains of NDP52 and OPTN are essential for the proteins' ability to translocate to depolarized mitochondria and for mitophagy to occur. Furthermore, deletion of TBK1, a kinase that

phosphorylates OPTN, revealed the importance of this phosphorylation for effective mitophagy.

Under the old model, Parkin-mediated ubiquitylation of OMM proteins is required to recruit autophagy receptors to mitochondria. Intriguingly, the authors found that attaching PINK1 to mitochondria recruited OPTN and NDP52, even in the absence of Parkin or mitochondrial damage. But the receptors could not translocate to mitochondria if ubiquitin was not phosphorylated, or if the receptors were mutated to lack their ubiquitin-binding domains. Finally, Lazarou and colleagues showed that OPTN and NDP52 bind preferentially to phospho-ubiquitin rather than to ubiquitin.

Armed with these data, the complex interactions at play can be untangled. After damage to the mitochondria, PINK1 phosphorylates ubiquitin that is already attached to OMM proteins. The resulting phospho-ubiquitin is crucial for recruiting OPTN and NDP52 to mitochondria. Parkin assists in the process, amplifying the phospho-ubiquitin by setting up a positive feed-forward cycle, so that more ubiquitin attaches to the OMM proteins. This enables OPTN and NDP52 to engage components of the autophagy machinery to initiate mitophagy (Fig. 1b).

This updated model is consistent with the authors' surprising observation that OPTN and NDP52 can be recruited to mitochondria independently of Parkin. However, the model does not imply that Parkin has no part to play, because Lazarou and colleagues found that Parkin expression dramatically increased

OPTN-dependent mitophagy. Strong genetic interactions between the PINK1 and PARKIN genes in model systems have been reported¹⁷. Moreover, familial Parkinson's disease can be caused by mutations in either PINK1 or PARKIN in humans¹. So Parkin clearly cooperates with PINK1 and has an important role in PINK1-mediated mitophagy under physiological conditions.

The findings of the current study involved only *in vitro* and cell-culture experiments, so their physiological significance *in vivo* remains to be demonstrated. Further biochemical and structural studies should explore how the phosphorylation of ubiquitin mediates its interactions with OPTN and NDP52. Such types of investigation are already improving our understanding of how Parkin is regulated by PINK1-phosphorylated ubiquitin — an interaction that was also discovered through *in vitro* and cell-culture experiments^{9–11}. For instance, a structural study¹⁸ published on Nature's website reveals the molecular basis for the way phospho-ubiquitin regulates Parkin. And a recent study¹⁹ of post-mortem brain samples and cells called fibroblasts from people with Parkinson's disease points to the physiological significance of phosphorylated ubiquitin in humans. Once Lazarou and colleagues' work has been similarly validated, it could establish a new way of thinking about how simple ubiquitin tags play multiple specific parts, and will surely provide inspiration for further exploration of ubiquitin and autophagy. ■

Noriyuki Matsuda is at the Ubiquitin Project,

Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan. Keiji Tanaka is at the Laboratory of Protein Metabolism, Tokyo Metropolitan Institute of Medical Science.

e-mails: matsuda-nr@igakuken.or.jp; tanaka-kj@igakuken.or.jp

1. Narendra, D., Walker, J. E. & Youle, R. *Cold Spring Harb. Perspect. Biol.* **4**, a011338 (2012).
2. Lazarou, M. *et al.* *Nature* **524**, 309–314 (2015).
3. Wild, P. & Dikic, I. *Nature Cell Biol.* **12**, 104–106 (2010).
4. Yoshii, S. R., Kishi, C., Ishihara, N. & Mizushima, N. *J. Biol. Chem.* **286**, 19630–19640 (2011).
5. Chan, N. C. *et al.* *Hum. Mol. Genet.* **20**, 1726–1737 (2011).
6. Padman, B. S., Bach, M., Lucarelli, G., Prescott, M. & Ramm, G. *Autophagy* **9**, 1862–1875 (2013).
7. Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. *Autophagy* **6**, 1090–1106 (2010).
8. Okatsu, K. *et al.* *Genes Cells* **15**, 887–900 (2010).
9. Kane, L. A. *et al.* *J. Cell Biol.* **205**, 143–153 (2014).
10. Kazlauskaitė, A. *et al.* *Biochem. J.* **460**, 127–139 (2014).
11. Koyano, F. *et al.* *Nature* **510**, 162–166 (2014).
12. Okatsu, K. *et al.* *J. Cell Biol.* **209**, 111–128 (2015).
13. Wauer, T. *et al.* *EMBO J.* **34**, 307–325 (2015).
14. Shiba-Fukushima, K. *et al.* *PLoS Genet.* **10**, e1004861 (2014).
15. Ordureau, A. *et al.* *Mol. Cell* **56**, 360–375 (2014).
16. Wong, Y. C. & Holzbaur, E. L. F. *Proc. Natl Acad. Sci. USA* **111**, E4439–E4448 (2014).
17. Pallanck, L. & Greenamyre, J. T. *Nature* **441**, 1058 (2006).
18. Wauer, T., Simicek, M., Schubert, A. & Komander, D. *Nature* <http://dx.doi.org/10.1038/nature14879> (2015).
19. Fiesel, F. C. *et al.* *EMBO Rep.* <http://dx.doi.org/10.15252/embr.201540514> (2015).

This article was published online on 12 August 2015.

DENGUE VIRUS

Bumps in the road to therapeutic antibodies

A human antibody against dengue virus serotype 2 has been shown to protect mice against disease. Structures of the antibody bound to the virus illuminate how it binds different viral forms to prevent virus entry into cells.

LESLIE GOO & THEODORE C. PIERSON

Dengue viruses are mosquito-borne flaviviruses that circulate as four related serotypes. Around 400 million human dengue virus infections occur across the globe annually. Roughly 25% of these result in clinical symptoms ranging from a self-limiting fever to a potentially fatal vascular-leakage disease¹. Despite this disease burden and decades of research, no effective vaccine or therapeutic agent is available. To redress this, much research has focused on understanding how antibodies bind to dengue virus particles and

inhibit infection, because the presence of neutralizing antibodies is thought to correlate with protection following vaccination, and such antibodies hold promise as antiviral drugs². Writing in *Science*, Fibriansah *et al.*³ show that the monoclonal antibody 2D22 binds multiple envelope proteins on the surface of the virus particle in a manner that prevents the conformational rearrangements necessary for virus entry into cells.

The viral envelope (E) protein is an elongated three-domain glycoprotein that orchestrates multiple steps in the virus-entry process, including attachment to cellular receptors and fusion

of viral and cellular membranes following internalization. E proteins are the principal targets of neutralizing antibodies, which may inhibit infection through several mechanisms. How antibodies engage E proteins on virus particles (virions), and in what numbers, can be influenced by the proteins' arrangement.

Pioneering cryo-electron microscopy analysis⁴ of dengue virions revealed a dense herringbone arrangement of E proteins comprising 30 rafts of three antiparallel dimers oriented flat against the viral membrane. Subsequent studies^{5,6} revealed that not all virions have this 'smooth' appearance and that, at 37 °C, the virions of some dengue strains take on an expanded structure with a 'bumpy' surface, on which the E-protein dimers are more loosely arranged (Fig. 1a). Whether antibodies distinguish between these smooth and bumpy structures, or influence transitions between these states, has not been extensively studied.

The antibody 2D22 was isolated from an individual with immunity to dengue virus serotype 2 (DENV2). This antibody neutralizes DENV2 (but not other serotypes) *in vitro*, and has been hypothesized to bind a molecular structure (epitope) spanning more than one

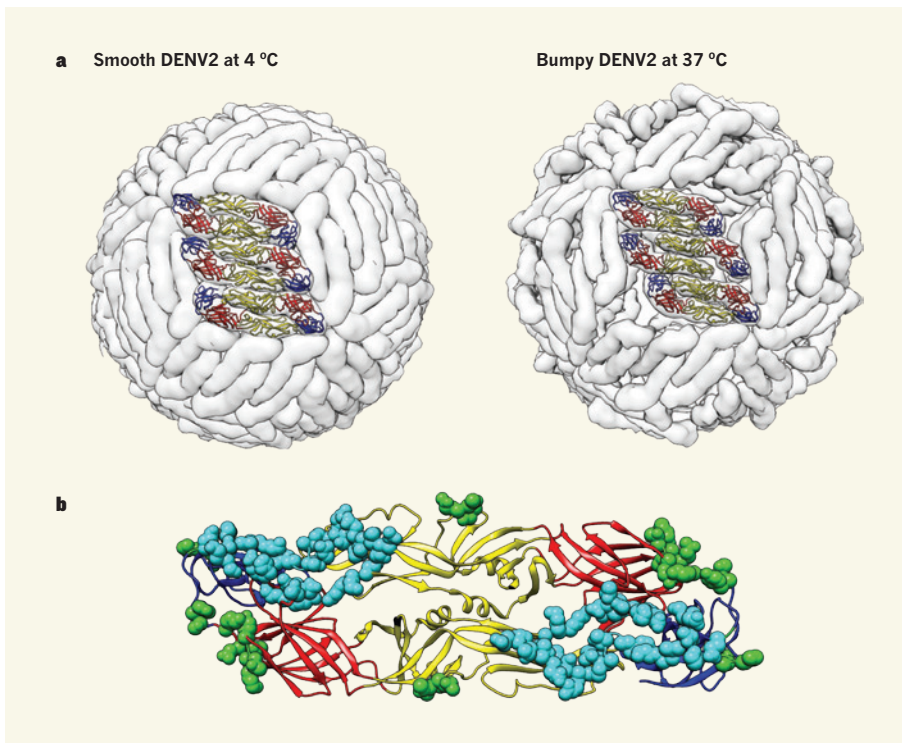


Figure 1 | Surface structure affects antibody binding. **a**, The virus particles of some strains of dengue virus serotype 2 (DENV2) adopt different surface conformations at different temperatures. At 4 °C, the surface-envelope (E) proteins are densely arranged as a 'smooth' configuration of 30 rafts of three antiparallel dimers (one raft of three dimers is coloured; the red, yellow and blue indicate the three domains of each monomer). At 37 °C, the surface becomes 'bumpy', with the E-protein dimers more loosely arranged. **b**, A close-up of a single E-protein dimer shows the molecular regions (epitopes) identified by Fibriansah *et al.*³ as being bound by the monoclonal antibody 2D22. When the virus surface is smooth, 2D22 binds to epitopes that span E proteins within a dimer (amino-acid residues represented as cyan spheres), and some that span adjacent dimers (green spheres). The latter contacts are predicted to be disrupted on bumpy virions. Molecular graphics and analyses were performed with the UCSF Chimera package. PDB IDs for smooth and bumpy virions are 4CCT and 3ZKO.

E protein⁷. Fibriansah *et al.* now demonstrate that 2D22 can protect against DENV2 infection in immunodeficient mice when administered before or after virus infection.

There is epidemiological evidence for increased risk of severe disease following secondary infection of an individual with a dengue serotype distinct from that of the primary infection⁸; this phenomenon has been linked to the activities of poorly neutralizing antibodies that bind more than one serotype, which initiate a mechanism called antibody-dependent enhancement of infection⁹. Fibriansah *et al.* show that variants of 2D22 engineered to be incapable of binding the Fc receptors present on many immune-system cells prevented the development of lethal antibody-mediated enhanced disease in immunodeficient mice.

The authors then used cryo-electron microscopy to solve the structures of 2D22 in complex with DENV2, for multiple DENV strains and at different temperatures. The structures show that the antibody binds an epitope that bridges E proteins within a dimer (Fig. 1b). This is similar to the recently reported mode of binding of a group of

potently neutralizing monoclonal antibodies that bind multiple dengue serotypes^{10,11}. However, although this E-dimer epitope (EDE) is located within E-protein dimers that are the 'building blocks' of the virions, it was not uniformly displayed on different DENV2 strains. The authors' analysis of a primary DENV2 viral strain revealed smooth virus particles at both 4 °C and 37 °C that were bound by 180 copies of 2D22. By contrast, the stoichiometry of antibody binding to the laboratory-adapted strain NGC was temperature dependent — when antibodies were bound to smooth virions at 4 °C and then the temperature was shifted to 37 °C, at which the virions adopt the bumpy configuration, one third of 2D22 epitopes became unavailable for binding.

This finding implies that the epitopes recognized by antibodies, including those within the E dimer, may not be equivalently accessible on bumpy and smooth particles, highlighting a need to understand the structural interactions that determine whether a given dengue strain will adopt an expanded or smooth structure at 37 °C. Notably, the 120 2D22 epitopes present on the NGC strain

at 37 °C is much greater than the number of antibodies predicted to be required for antibody-mediated neutralization of flaviviruses¹².

Multiple lines of evidence suggest that the structure of dengue viruses can be variable at physiological temperatures. Not all structures observed by microscopy are represented by the smooth and bumpy configurations, owing in part to inefficient maturation or conformational flexibility of the virion¹³. How does this complexity affect antibody recognition and immunity? Researchers have characterized several potentially neutralizing human antibodies that bind complex epitopes made up of more than one E protein^{14,15}, including others that bind to EDE^{10,11}. This form of recognition should in principle be sensitive to changes in the surface of the virion that alter the herringbone arrangement of E proteins, as illustrated by Fibriansah and colleagues' data. Amino-acid variation that alters higher-order E-protein structures or the strength of dimer contacts thus has the potential to affect the neutralization potency of these antibodies.

Conversely, antibodies may have an active role in shaping the surface of infectious virions. Some EDE-binding antibodies can 'push' immature virus structures into the E-dimer configuration found on mature virions, allowing these antibodies to neutralize the complex spectrum of virus geometries released from infected cells^{10,11}. The identification of potentially neutralizing human antibodies such as 2D22 that bind complex epitopes is an important step toward identifying targets for vaccine and drug design. Such antibodies also have the potential to provide useful insight into the many shapes and forms of infectious virions. ■

Leslie Goo and Theodore C. Pierson are in the Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: piersontc@mail.nih.gov

1. Bhatt, S. *et al.* *Nature* **496**, 504–507 (2013).
2. Murphy, B. R. & Whitehead S. S. *Annu. Rev. Immunol.* **29**, 587–619 (2011).
3. Fibriansah, G. *et al.* *Science* **349**, 88–91 (2015).
4. Kuhn, R. J. *et al.* *Cell* **108**, 717–725 (2002).
5. Fibriansah, G. *et al.* *J. Virol.* **87**, 7585–7592 (2013).
6. Zhang, X. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 6795–6799 (2013).
7. de Alwis, R. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 7439–7444 (2012).
8. Burke, D. S., Nisalak, A., Johnson, D. E. & Scott, R. M. *Am. J. Trop. Med. Hyg.* **38**, 172–180 (1988).
9. Halstead, S. B. *Adv. Virus Res.* **60**, 421–467 (2003).
10. Dejnirattisai, W. *et al.* *Nature Immunol.* **16**, 170–177 (2015).
11. Rouvinski, A. *et al.* *Nature* **520**, 109–113 (2015).
12. Dowd, K. A. & Pierson, T. C. *Virology* **411**, 306–315 (2011).
13. Kuhn, R. J., Dowd, K. A., Post, C. B. & Pierson, T. C. *Virology* **479–480**, 508–517 (2015).
14. Teoh, E. P. *et al.* *Sci. Transl. Med.* **4**, 139ra83 (2012).
15. Kaufmann, B. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 18950–18955 (2010).

ORGANIC CHEMISTRY

Light opens pathways for nickel catalysis

The report of a light-activated catalyst that generates usually inaccessible nickel intermediates provides a general strategy that might allow elusive catalytic reactions to be realized. [SEE LETTER P.330](#)

JOHN J. MURPHY & PAOLO MELCHIORRE

For catalysis mediated by transition metals, the catalyst's oxidation state mainly dictates which reactions are possible. The ability to modulate the inherent oxidation state of a metal intermediate by applying an external stimulus can therefore enable otherwise inaccessible catalytic modes of reactivity, opening up unprecedented opportunities for innovation in metal-mediated catalysis. On page 330, Terrett *et al.*¹ describe how this ideal can be achieved using a catalyst activated by visible light. Their strategy allowed them to develop a general method of coupling carbon atoms to oxygen atoms by using a nickel catalyst, even though nickel is usually resistant to this type of chemical transformation.

The field of nickel catalysis has made tremendous advances in the past decade², with valuable chemical transformations being developed to synthesize natural products, polymers and pharmaceuticals. Nickel is the lightest element in group 10 of the periodic table, a group it shares with palladium and platinum, and it has similar reactivity to these elements. It can mediate reactions to construct carbon–carbon and carbon–heteroatom bonds (heteroatoms are any atoms other than carbon or hydrogen), complementing established protocols catalysed by more-expensive noble metals.

Nickel is particularly attractive for catalysis because it has several oxidation states that are primed for distinct modes of reactivity. Nickel-catalysed reactions generally involve redox events in which one or two electrons are transferred, and which interconvert organometallic nickel intermediates, changing the metal's oxidation state. In particular,

high-valent nickel complexes (in which nickel is in oxidation state +3 or +4) have a greater tendency to promote carbon–heteroatom coupling reactions^{3,4} than do their lower-valent counterparts (in which nickel is in oxidation states 0, +1 or +2). In chemistry jargon, they are more susceptible to reductive elimination.

For example, organonickel(II) alkoxides are complexes in which nickel in oxidation state

+2 is bound to both a carbon and an oxygen atom. These complexes are easily accessible from nickel(0) but do not readily undergo carbon–oxygen (C–O) bond-forming reactions. Instead, a nickel(III) intermediate, generated through single-electron-transfer oxidation, is required for efficient reductive elimination. This reactivity has been used in non-catalytic C–O bond-formation reactions^{3–5} in which preformed organo-nickel(II) alkoxides are oxidized. But developing a catalytic nickel-mediated system to forge C–O bonds⁶ poses more of a conundrum. The main problem is the need for an oxidant that not only promotes the formation of the crucial high-valent intermediates (organo-nickel(III) alkoxides) from catalytic nickel(II) precursors, but is also compatible with the reactivity of metal species in lower oxidation states — mainly nickel(0) species — that are also present in the catalytic cycle.

Terrett and colleagues have solved this problem by integrating the redox activity of an iridium(III) catalyst that absorbs visible light (an iridium(III) photocatalyst) with a nickel catalytic cycle. They used this strategy to prepare a wide array of synthetically valuable compounds called alkyl aryl ethers at room temperature from readily available starting materials (alcohols and aryl bromides; Fig. 1). Crucially, on absorption of weak light, the iridium(III) photocatalyst enters an excited state ($^*Ir^{III}$) in which iridium can readily engage in an array of redox processes. Specifically, $^*Ir^{III}$ takes an electron from the otherwise inert organonickel(II) alkoxide to yield an electron-rich iridium(II) centre and the requisite nickel(III) alkoxide, which then undergoes the desired C–O coupling to form the alkyl aryl ether.

After the product has been released, the nickel catalyst, now in oxidation state +1, must be reactivated so that it can undergo further catalytic cycles. The electron-rich iridium(II) species gives it an electron to form a nickel(0) species, which can easily add to an aromatic bromide (an event known as oxidative addition). This final electron-transfer process simultaneously returns the two metal catalysts to their original oxidation states.

Terrett and co-workers' proposed mechanism suggests that multiple nickel species, in different oxidation states, are transiently generated. Therefore, the problem that typically complicates

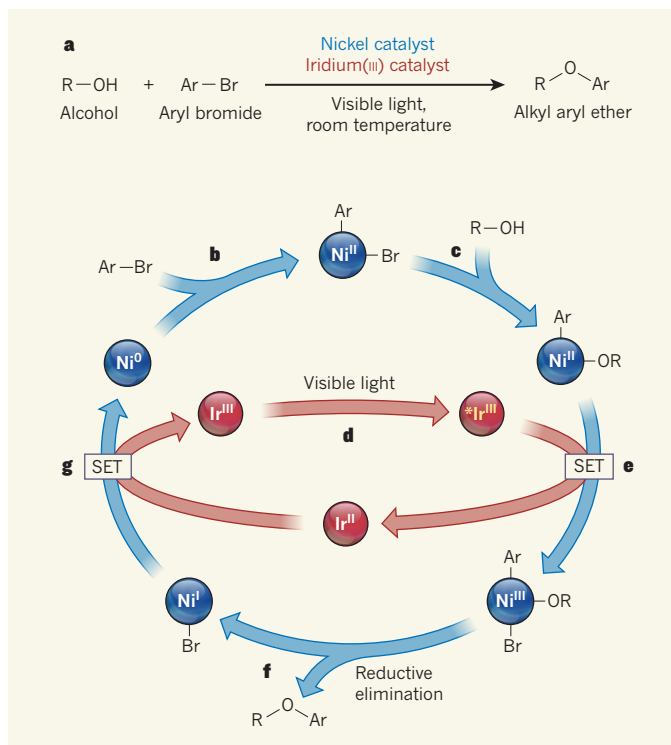


Figure 1 | Switching oxidation states for nickel catalysis. **a**, Terrett *et al.*¹ report a method for constructing alkyl aryl ethers from simple alcohols and aryl bromides at room temperature, in which the redox activity of a light-activated iridium(III) catalyst is integrated with a nickel catalytic cycle. R represents different attached chemical groups; Ar represents aromatic groups (benzene rings and related structures); Br, bromine. **b**, In the proposed mechanism, the aryl bromide adds to a nickel(0) species, producing a nickel(II) intermediate, $ArNi(II)Br$. **c**, This reacts with the alcohol to generate an inert organo-nickel(II) alkoxide, $ArNi(II)OR$. **d, e**, The authors propose that when the iridium(III) catalyst is irradiated by light (**d**), the resulting excited complex ($^*Ir^{III}$) oxidizes $ArNi(II)OR$ through single-electron transfer (SET; **e**), generating an Ir(II) species and a nickel(III) intermediate, $ArNi(III)ORBr$. **f**, This intermediate undergoes reductive elimination to yield an alkyl aryl ether and a nickel(I) species, $Ni(I)Br$. **g**, SET from the iridium(II) species to $Ni(I)Br$ restores the original oxidation states of the two metal catalysts, completing both catalytic cycles.

the development of nickel-catalysed processes under oxidizing conditions — achieving oxidation of the nickel(II) intermediate without oxidizing the nickel(0) species — might also be an issue here. But because the iridium photocatalyst is partitioned within different oxidation states by light activation, it can donate and receive electrons at different stages of the reaction, ensuring a kind of redox equilibrium. So although nickel(0) oxidation could occur through the action of the iridium photocatalyst, this would probably be a reversible process; only when the crucial organo-nickel(III) alkoxide is generated will productive C–O coupling occur.

Over the past few years, the redox properties of iridium photocatalysts have been used to develop a variety of light-driven transformations, establishing a field known as photoredox catalysis⁷. Combinations of photoredox and nickel catalysis are also known^{8,9}, in which the photocatalytic process generates carbon-centred radicals, which are then captured by nickel(II) intermediates. But Terrett and co-workers' study is particularly exciting because it demonstrates that the light-activated iridium catalyst can participate directly in the nickel catalytic mechanism by leveraging the oxidation state of the organometallic intermediates.

Moreover, the authors provide a great practical advance in the catalytic formation of alkyl aryl ethers. The creation of such ethers from alcohols and aryl halides is one of the most widely used chemical transformations in modern synthetic chemistry. But Terrett and colleagues' method requires much milder reaction conditions than do existing catalytic technologies, including those based on palladium¹⁰ or copper¹¹ — that is, the reactions can be conducted at ambient temperature by using blue-light-emitting diodes as the energy source, rather than at higher temperatures, and avoid the use of strong bases that can compromise the stability of the reagents. What is more, water can be used as a substrate instead of alcohols, allowing easy access to products called phenols. The authors demonstrate a wide range of substrates, but there are still many opportunities to expand the scope of the reaction.

The discovery that a photoredox catalyst, activated by weak light, can unlock previously inaccessible nickel coupling mechanisms will greatly interest synthetic organic chemists, and create opportunities to realize other elusive catalytic reactions. In a similar vein, a ruthenium photocatalyst has recently been reported¹² to enable a difficult carbon–nitrogen bond-forming reaction by driving the oxidative formation of a key nickel(III) intermediate. Further insights into the intricate networks of redox processes in these related reactions are needed to translate the methodological blueprint to other metal species and organometallic reactivities. ■

John J. Murphy and Paolo Melchiorre are at

the Institute of Chemical Research of Catalonia (ICIQ), Tarragona 43007, Spain. P.M. is also at the Catalan Institution for Research and Advanced Studies, Barcelona, Spain. e-mail: pmelchiorre@iciq.es

1. Terrett, J. A., Cuthbertson, J. D., Shurtleff, V. W. & MacMillan, D. W. C. *Nature* **524**, 330–334 (2015).
2. Tasker, S. Z., Standley, E. A. & Jamison, T. F. *Nature* **509**, 299–309 (2014).
3. Camasso, N. M. & Sanford, M. S. *Science* **347**, 1218–1220 (2015).
4. Zhou, W., Schultz, J. W., Rath, N. P. & Mirica, L. M. *J. Am. Chem. Soc.* **137**, 7604–7607 (2015).

5. Han, R. & Hillhouse, G. L. *J. Am. Chem. Soc.* **119**, 8135–8136 (1997).
6. Mann, G. & Hartwig, J. F. *J. Org. Chem.* **62**, 5413–5418 (1997).
7. Prier, C. K. *et al.* *Chem. Rev.* **113**, 5322–5363 (2013).
8. Tellis, J. C., Primer, D. N. & Molander, G. A. *Science* **345**, 433–436 (2014).
9. Zuo, Z. *et al.* *Science* **345**, 437–440 (2014).
10. Torracca, K. E., Huang, X., Parrish, C. A. & Buchwald, S. L. *J. Am. Chem. Soc.* **123**, 10770–10771 (2001).
11. Wolter, M. *et al.* *Org. Lett.* **4**, 973–976 (2002).
12. Tasker, S. Z. & Jamison, T. F. *J. Am. Chem. Soc.* **137**, 9531–9534 (2015).

This article was published online on 12 August 2015.

STRUCTURAL BIOLOGY

Hypoxia response becomes crystal clear

The crystal structures of two proteins that respond to reduced tissue oxygen levels — hypoxia-inducible factors — provide insight into their function and reveal sites for rational drug design. [SEE ARTICLE P.303](#)

RONEN MARMORSTEIN & M. CELESTE SIMON

Oxygen is a key substrate for cellular energy production and numerous intracellular biochemical reactions. Maintaining oxygen homeostasis is therefore essential for the survival of most species. Oxygen deprivation (hypoxia) triggers complex responses at the cellular, tissue and organismal levels, to match oxygen supply with metabolic and bioenergetic demands. Several evolutionarily conserved molecular responses to hypoxia are mediated by hypoxia-inducible factors¹ (HIFs) — transcription factors that regulate metabolism, stem-cell behaviour, the generation of blood vessels and red blood cells, inflammatory responses and cell motility. The first crystal structures of two HIFs are described by Wu *et al.*² on page 303 of this issue.

The involvement of HIFs in so many cellular pathways means that the proteins play a central part in normal development and physiology, and are dysregulated in many diseases, including cancer. Yet despite the fascinating biology that they mediate, and their importance as drug targets, the molecular basis for HIF activity has remained incompletely characterized, impeding the development of highly specific inhibitors. A primary reason for this is that HIF proteins, which are heterodimeric complexes consisting of subunits of HIF- α and ARNT (aryl hydrocarbon receptor nuclear translocator), have not previously been crystallized for high-resolution X-ray structure determination (other than heterodimers of the PAS-B domains of HIF-2 α and ARNT; ref. 3). Wu *et al.* have now achieved this, and present

structures of HIF-1 α –ARNT and HIF-2 α –ARNT complexes — alone, bound to their corresponding hypoxia-response DNA elements (HREs), and bound to the previously identified small-molecule inhibitors OX3 (ref. 4) and proflavine⁵.

The structures show that the HIF-2 α and ARNT subunits wrap around each other in a non-symmetrical way, such that their respective basic helix–loop–helix (bHLH), PAS-A and PAS-B domains make different contacts with neighbouring domains (Fig. 1a). Although the domains of HIF-2 α interact contiguously with each other to generate an elongated substructure, the ARNT domains do not interact with each other, but instead bind HIF-2 α domains in the heterodimeric complex. This form of dimerization is highly unusual. Moreover, the domain–domain interactions are distinct from those seen in another heterodimeric transcription factor that has bHLH and PAS domains, the CLOCK–BMAL1 complex⁶. These observations suggest that ARNT is highly selective for associations with HIF-1 α and HIF-2 α , and that the protein subunits also help to stabilize the complex. Wu *et al.* confirmed this function by mutating ARNT and HIF- α interface amino-acid residues and showing that several of these mutations inhibit association of the subunits *in vivo*.

The bHLH–PAS domain structure is characteristic of several related family-member transcription factors, such as AhR (ref. 7) and NPAS2 (ref. 8), which have been shown to use their respective PAS motifs to bind small molecules. But whether the same is true for HIF complexes was not known. Wu and colleagues'

structure of HIF-2 α -ARNT reveals five potential binding 'pockets' for small-molecule compounds, one in each of the PAS domains of HIF-2 α and ARNT, with the fifth lying at the interface between HIF-2 α PAS-B and ARNT PAS-A (Fig. 1a). The authors also present a structure of HIF-2 α -ARNT bound to the small molecule proflavine indicating that it binds to the interface pocket. This is consistent with the previous observation that acriflavine (a mixture of proflavine and tryptaflavin) destabilizes HIF-2 α -ARNT heterodimers in cells⁵. A further structure reveals that the HIF-2 α inhibitor OX3 (ref. 4) binds to the pocket on the HIF-2 α PAS-B domain.

The authors also compared the amino-acid sequences of the binding pockets of HIF-1 α and HIF-2 α , and found the interface (proflavine-binding) pocket to be highly conserved, but the PAS-A and PAS-B pockets less so. This suggests that small-molecule inhibitors may be developed that are selective for complexes containing either HIF-1 α or HIF-2 α , or that target both complexes. These are both highly desirable goals for drug design, depending on the disease context⁹. The findings also reinvigorate the possibility that HIF proteins bind their cellular ligands through one or more of these subdomains, as has been shown for AhR-ligand binding in the gastrointestinal system¹⁰.

Wu and colleagues went on to determine the structure of both HIF-1 α -ARNT and HIF-2 α -ARNT bound to an HRE sequence — the first report of a multidomain bHLH-PAS heterodimer in association with DNA. As expected, both HIF- α -ARNT dimers in complex with DNA show the same overall architecture, and

they use conserved amino-acid residues in a DNA-recognition helix region (α 1) of the bHLH domain to make base-specific contacts in the major groove of the DNA. However, unexpectedly, a comparison of the DNA-free and DNA-bound complexes revealed that the amino-terminal end of the α 1 helix undergoes a coil-to-helix transition that extends the helix to facilitate the base-specific interactions (Fig. 1b). The preceding bHLH α 2 helix also rotates to help orient the α 1 helix for DNA recognition. Another surprise from the structure is that the PAS-A domain of each HIF- α subunit makes contact with the minor groove of the DNA approximately six base pairs away from the HRE core element, thus extending the 'footprint' of the HIF- α -ARNT heterodimer on DNA. These observations reveal that HIF- α -ARNT complexes undergo structural alteration of the bHLH region, with participation of the HIF- α PAS-A domain, for DNA recognition.

Mutations in both HIF-1 α and HIF-2 α have been identified from cancer biopsies^{11,12}, and Wu *et al.* show that these mutations correspond to amino-acid residues at interfaces between protein domains, DNA-contact residues or pocket regions in the PAS domains. This finding reinforces the biological importance of crystallographic information on the HIF- α -ARNT complexes, their mode of DNA recognition and the binding of small molecules.

Finally, Wu and colleagues' findings suggest that different bHLH-PAS family members integrate distinct environmental cues, and that there is not one mode of heterodimerization for the bHLH-PAS proteins. Such differently arranged domain-domain junctions

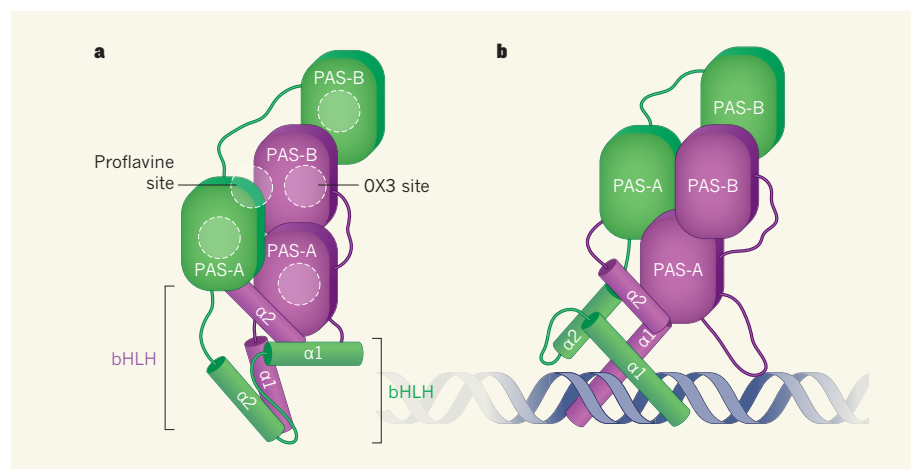


Figure 1 | The HIF- α -ARNT heterodimer off and on DNA. **a**, The two protein subunits of the transcription factor hypoxia-inducible factor 2 (HIF-2) — HIF-2 α (purple) and ARNT (green) — each have bHLH, PAS-A and PAS-B domains. Wu and colleagues' crystal structure² of this heterodimer reveals contiguous interactions between the HIF-2 α domains, but not between those of ARNT. Furthermore, the structures show binding pockets for putative small-molecule ligands (circles), including proflavine and OX3. **b**, The authors also present a structure of the heterodimer in complex with DNA. The domain organization seen in this structure highlights the extension of the bHLH α 1 helices such that they can interact with the DNA major groove; reorient the α 2 helices to position the α 1 helices for DNA recognition; and direct minor-groove DNA interactions with the PAS-A domain of HIF-2 α .



50 Years Ago

A Star Called the Sun by Dr. George Gamow — We are reminded of Dante on his journey to the inferior regions: "I would have cast me into molten glass to cool me, when I enter'd, so intense raged the conflagrant mass" ... the author compares the Sun with other stars and explores the nature of such old ones as red giants, white dwarfs and supernovae, the stars that "die gloriously" ... The style is popular and sometimes racy. Occasionally it will offend the English reader who has some feeling for his Mother tongue, but on the whole, in spite of his "oodles and oodles", the author has avoided ... ambiguities and half-statements. **W. L. Summer**
From *Nature* 21 August 1965

100 Years Ago

The Book of France edited by Winifred Stephens — The book opens with an appreciation of France by Mr. Henry James, written with the copiousness and *verve* which characterize all his work ... Perhaps the most arresting and striking contribution to the work is the "Debout pour la Dernière Guerre!" ... done into nervous, palpitating English by Mr. H. G. Wells ... "La Basilique-Fantôme", by Pierre Loti ... is a remarkable word-picture of the present state of the cathedral of Rheims ... "The most irreparable disaster is that of those great stained windows composed by the mysterious artists of the thirteenth century ... Masterpieces that no one can reproduce showered down their fragments never to be sorted again, their wonderful golds and reds and blues, of which the secret has been lost, upon the pavement stones. Gone for ever those rainbow transparencies, gone for ever those companies of saints with the charm of their simple attitudes." **T. E. Thorpe**
From *Nature* 19 August 1915

not only promote stability, but also allow the proteins to incorporate diverse signals and communicate these signals across domains. These signals may include cellular ligands, disease-related mutations and post-translational modifications. The authors' solving of the HIF- α -ARNT structures provides exciting opportunities to explore previously uncharacterized cell-regulatory pathways and to advance drug discovery. ■

Ronen Marmorstein and M. Celeste Simon are in the Abramson Family Cancer Research

Institute, and Departments of Biochemistry and Biophysics (R.M.) and Cell and Developmental Biology (M.C.S.), University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA.

e-mails: marmor@mail.med.upenn.edu; celeste2@mail.med.upenn.edu.

1. Semenza, G. L. *Cell* **148**, 399–408 (2012).
2. Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. *Nature* **524**, 303–308 (2015).
3. Scheuermann, T. H. et al. *Proc. Natl Acad. Sci. USA* **106**, 450–455 (2009).

4. Scheuermann, T. H. et al. *Nature Chem. Biol.* **9**, 271–276 (2013).
5. Lee, K. et al. *Proc. Natl Acad. Sci. USA* **106**, 17910–17915 (2009).
6. Huang, N. et al. *Science* **337**, 189–194 (2012).
7. Denison, M. S., Soshilov, A. A., He, G., DeGroot, D. E. & Zhao, B. *Toxicol. Sci.* **124**, 1–22 (2011).
8. Dioum, E. M. et al. *Science* **298**, 2385–2387 (2002).
9. Keith, B., Johnson, R. S. & Simon, M. C. *Nature Rev. Cancer* **12**, 9–22 (2012).
10. Li, Y. et al. *Cell* **147**, 629–640 (2011).
11. Ollerenshaw, M., Page, T., Hammonds, J. & Demaine, A. *Cancer Genet. Cytogenet.* **153**, 122–126 (2004).
12. Morris, M. R. et al. *Anticancer Res.* **29**, 4337–4343 (2009).

This article was published online on 5 August 2015.

BIODIVERSITY

Multiple origins of mountain life

A study of DNA sequences from more than 1,800 organisms on Mount Kinabalu in Borneo reveals the evolutionary mechanisms that led to the mountain's high and unique biodiversity. [SEE LETTER P.347](#)

ALEXANDRE ANTONELLI

Mountains occupy only about one-eighth of the world's land surface outside Antarctica, yet they are home to around one-third of all terrestrial species¹. However, little is known about when, where and how this mountain biodiversity developed. In a paper on page 347, Merckx *et al.*² investigate the origins of species on a highly diverse tropical mountain, Mount Kinabalu on the island of Borneo. They find that most endemic species — those found nowhere else — arose relatively recently on the mountain, deriving from both local ancestors and distant

immigrants that were pre-adapted to cool environments. These observations have direct implications for our understanding of montane biodiversity, and may offer clues on how to protect both the unique organisms found there and the habitats that provide the stage for speciation.

Ever since the first botanical documentation of a tropical mountain by Alexander von Humboldt more than two centuries ago³, naturalists have been fascinated by the diverse and unusual variety of life forms found on mountains. But we still lack answers to fundamental questions about the evolution of mountain biodiversity. Did the species living

on mountains originate in the surrounding lowlands, where their ancestors became successively adapted to higher altitudes? Or are montane species mostly pre-adapted immigrants from far-away mountains? And are the mountain-dwelling species ancient or recent?

Merckx and colleagues got to the root of these questions by using a new comparative approach. Rather than studying the formation of species in a single group of organisms and extrapolating from there, which has been the standard approach so far, the researchers collected a large variety of organisms — from frogs and snails to insects, plants and fungi — that inhabit the iconic Mount Kinabalu and its surroundings. They then sequenced DNA from the approximately 1,850 collected specimens, compared these sequences with others in their own collections and public databases, and calculated their relationships, ages, geographical origins and ancestral environments.

The first striking result is that most of the montane organisms examined are relatively young. They started to speciate during the past 6 million years, after, or at the same time as, the rise of the mountain they inhabit. Unlike the ancient creatures found on a remote tropical mountain in Arthur Conan Doyle's novel *The Lost World*, this finding suggests a recent origin

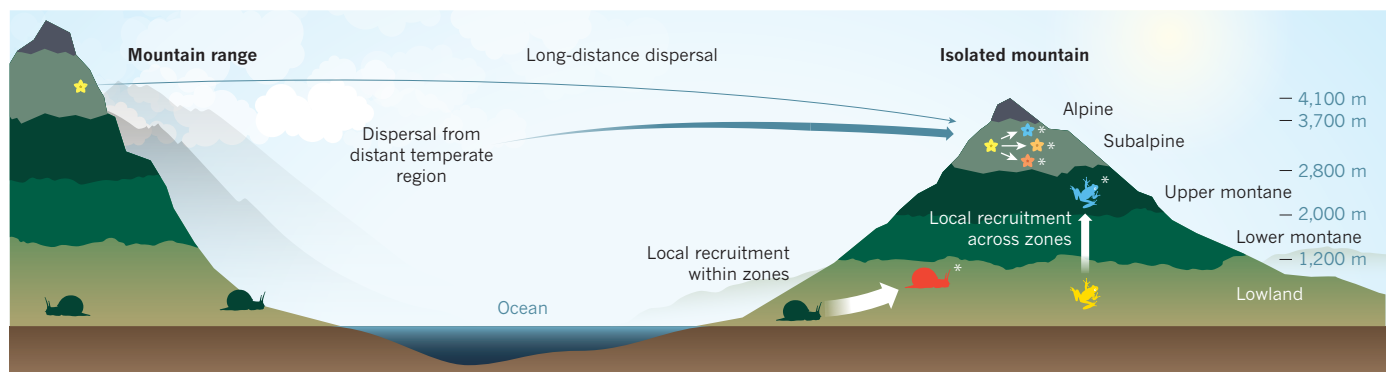


Figure 1 | Routes to mountain biodiversity. Isolated tropical mountains generally contain high levels of species richness and endemism (species uniqueness). Merckx *et al.*² show that a large proportion of endemic species (indicated by asterisks) on Mount Kinabalu in Borneo derive from lineages that were previously present on the island (arrow widths reflect the relative number of identified cases; organisms depicted are indicative only). Only some of these locally recruited species have adapted to different vegetation zones. By contrast, some species, especially those found at high altitudes, have

their origins in similar vegetation zones on other mountain ranges outside Borneo, or at lower altitudes in temperate regions, and have arrived by means of long-distance dispersal. (Note that vegetation zones on mountain ranges are usually at higher elevations than on isolated mountains, owing to heat retention and wind shadowing.) Most of these immigrant lineages have then undergone local speciation. The evolutionary history of Mount Kinabalu's biodiversity exemplifies the interactions among dispersal, adaptation and speciation in generating mountain biodiversity.

for montane species across the domains of life, and supports the recent speciation documented for alpine plants on several continents⁴.

The second major finding is the dual origin of montane organisms (Fig. 1). Some of the species, in particular those found at the highest elevations of Mount Kinabalu, have their closest relatives outside of Borneo. Their ancestors were often good at dispersing, such as plants or fungi that produced large quantities of light seeds or spores that could be transported with the wind. Other species — about twice as many — derive from local ancestors at lower altitudes on the same island. The location of Mount Kinabalu, surrounded by an exceedingly diverse tropical forest at the crossroads of Asia and Oceania, two regions that have their own distinct fauna and flora, apparently provided the mountain with a rich initial stock for the evolution of its unique biodiversity.

Finally, the authors' analysis shows an overarching role for niche conservatism — the tendency for organisms to maintain their environmental preferences over evolutionary time. This result is evident both from the immigrant and the local lineages that gave rise to Mount Kinabalu's biodiversity. Most of the ancestral species were already adapted to cool conditions, either in temperate regions or in other montane habitats. Even the lineages that 'climbed up' Mount Kinabalu often remained in the same broadly defined vegetation zone. The niche conservatism and pre-adaptation shown for the inhabitants of this tropical mountain are in line with previous findings across the Southern Hemisphere⁵ and with observed patterns of plant movement into cold environments around the world⁶.

Merckx and colleagues' study thus provides a textbook example of how biodiversity originates from the interplay between long-distance dispersal and local recruitment, followed by adaptation and speciation through interaction with changes in the landscape, climate and environment⁷. Its limitations are shared with other studies that are based on living organisms and current species distributions. Biological surveys sample only a fraction of the total biodiversity of an ecosystem, and large organisms found in easily accessible sites are typically over-represented. Furthermore, estimates of speciation events, geographical history and niche conservatism obtained from phylogenetic trees are prone to large error intervals and many assumptions, and largely disregard the confounding effects of extinction. The integration of recent and past data — from DNA, fossils and environmental and geological proxies — could remedy these shortcomings⁸, but such data are still scarce or not readily available.

Much could be gained by applying Merckx and colleagues' whole-community approach to other systems around the world. As well as increasing our fundamental knowledge of the diversity and distribution patterns of species, such eco-evolutionary studies would shed light

on why some regions are much more biodiverse than others. A better understanding of the past may also, at least to some extent, help scientists to manage the present and predict the future. Taking climate changes as an example, evolutionary studies may help to inform us about the resilience of species and ecosystems⁹. They may also help to ascertain the role of mountains as potential reservoirs of biodiversity, because montane species need to move only short distances to keep their preferred niche¹⁰. Finally, we may need to devote more resources to preserving the natural corridors that link vegetation zones along altitudinal slopes if we are to safeguard the biotic interchange reported by Merckx *et al.*, and thus enhance our protection of the world's unique and rich mountain biodiversity. ■

Alexandre Antonelli is in the Department of Biological and Environmental Sciences,

University of Gothenburg, SE 405 30 Gothenburg, Sweden; and Gothenburg Botanical Garden, SE 413 19 Gothenburg, Sweden.

e-mail: alexandre.antonelli@bioenv.gu.se.

1. Spehn, E. M., Rudmann-Maurer, K. & Körner, C. *Plant Ecol. Divers.* **4**, 301–302 (2011).
2. Merckx, V. S. F. T. *et al. Nature* **524**, 347–350 (2015).
3. Humboldt, A. & Bonpland, A. *Essai sur la géographie des plantes* (Chez Levrault, Schoell, 1805).
4. Hughes, C. E. & Atchison, G. W. *New Phytol.* **207**, 275–282 (2015).
5. Crisp, M. D. *et al. Nature* **458**, 754–756 (2009).
6. Zanne, A. E. *et al. Nature* **506**, 89–92 (2014).
7. Hoon, C., Mosbrugger, V., Mulch, A. & Antonelli, A. *Nature Geosci.* **6**, 154 (2013).
8. Fritz, S. A. *et al. Trends Ecol. Evol.* **28**, 509–516 (2013).
9. Hoffmann, A. A. & Sgrò, C. M. *Nature* **470**, 479–485 (2011).
10. Sandel, B. *et al. Science* **334**, 660–664 (2011).

This article was published online on 12 August.

ASTROPHYSICS

Mystery survivor of a supermassive black hole

The G2 cloud in our Galaxy's core has survived an encounter with the central black hole and failed to trigger a major flare-up in the black hole's activity. A promising theory endeavours to explain the cloud's nature.

JOHN BALLY

The centre of the Milky Way hosts a supermassive black hole (SMBH) weighing approximately 3.6 million solar masses. In 2012 a faint, dusty object of cloudy appearance was discovered¹, accelerating towards the SMBH on an eccentric orbit with a predicted closest approach (periapse) of only about 200 times the mean Earth–Sun separation (Fig. 1). During its encounter with the SMBH, the object, known as G2, avoided complete tidal disruption by the black hole's gravity, implying that it is gravitationally bound (held together by gravity). But what is the nature of G2? Various early searches failed to find an ordinary or an evolved giant star lurking in the cloud. Writing in the *Astrophysical Journal*, Mapelli and Ripamonti² propose that the central object might be a planetary embryo that was dynamically ejected from its parent system and is nearly as large as the distance between Earth and the Sun.

The central few light years of the Milky Way encompass a dense stellar field containing more than 10 million low-mass stars (Fig. 2); a comparable volume near the Sun typically only contains one star. Among these dim, ancient stars, there is a swarm of luminous massive

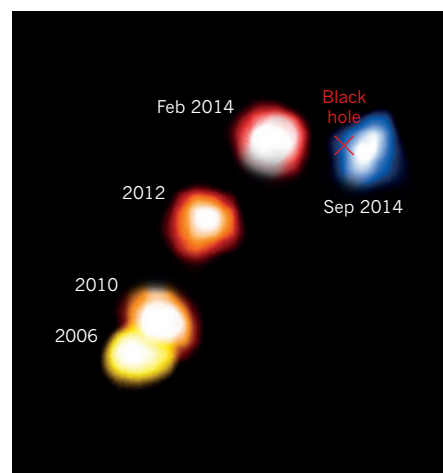


Figure 1 | At close quarters to Sagittarius A*.

The G2 cloud is shown at different times along an orbit that took it perilously close to Sagittarius A*, the supermassive black hole in our Galaxy's centre (red cross). These observations show that G2 was on a course towards the black hole from 2006 to early 2014 (reddish colours indicate that the object is receding from the observer). Its closest approach to Sagittarius A* was in May 2014 and observations in September 2014 (blue indicates that G2 is approaching) show that the object survived being ripped apart by the black hole.

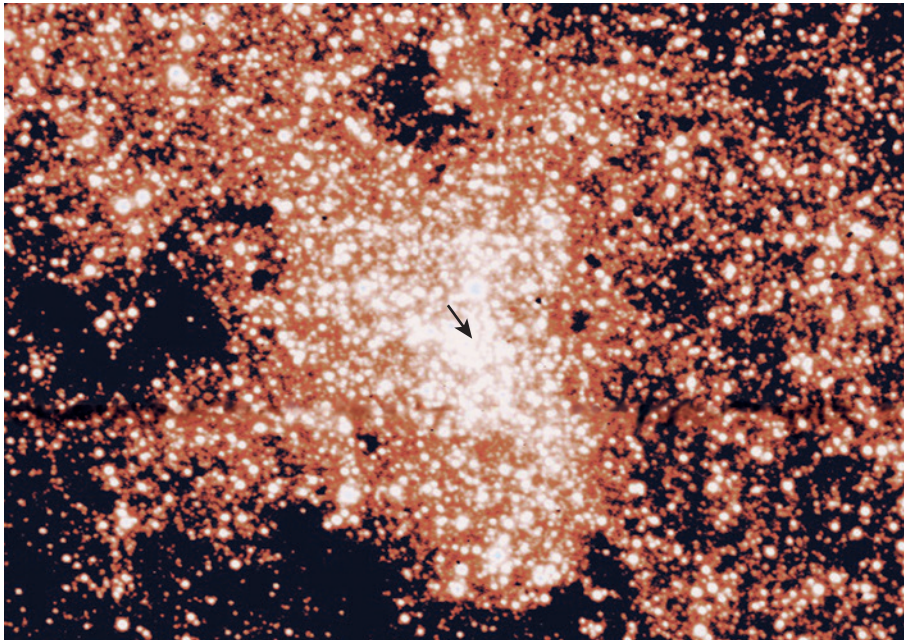


Figure 2 | The Galactic Centre. The numerical density of stars in the central few light years of our Galaxy is at least 10^7 times that in the Sun's neighbourhood. This high density can be glimpsed from this high-resolution, near-infrared view of the Galactic Centre, which shows young massive stars and old red giant stars. The location of our Galaxy's supermassive black hole, Sagittarius A*, is indicated. Mapelli and Ripamonti² propose that the object G2 may be a recently formed protoplanet that was previously ejected from its parent stellar system and was set on an orbit taking it close to Sagittarius A*.

stars formed in the past 10 million years³, surrounded by a ring of molecular gas containing the equivalent mass of tens of thousands of Suns. The young hot stars emit intense ultraviolet light that ionizes a region of hot plasma, which is heated by shock waves driven into it by powerful stellar winds. Right in the middle of this region lurks the SMBH known as Sagittarius A* (ref. 4).

The discovery of the G2 cloud's headlong plunge towards the central black hole caused much excitement among astronomers⁵. Models of G2's orbit predicted that the cloud would pass close to the black hole sometime during the spring of 2014. Even by late 2013, there was tentative evidence that parts of a tidal gas stream escaping from G2 had passed behind the SMBH and were emerging on the other side. Would the black hole swallow enough gas to ignite a burst of emissions activity? Would our Galaxy's centre suddenly brighten across the spectrum from X-rays to radio wavelengths and briefly resemble the powerful active galactic nuclei (quasars) associated with SMBHs in some galaxies? Dozens of papers emerged with predictions. Numerical models⁶ of clouds on plunging orbits predicted that a considerable amount of gas would spiral into the clutches of Sagittarius A*, igniting fireworks.

A year after the plunge, we can conclusively state that there was no major mass-accretion event onto the SMBH. Periape of G2 occurred in May 2014 and there was no subsequent flare-up of Sagittarius A*, which continued its dim flickering in X-ray and near-infrared wavelengths, and remained completely

oblivious to the object zipping by at thousands of kilometres per second.

What created the gas and dust seen as G2? An obvious candidate was the swollen atmosphere of a red giant or supergiant star. But such stars are luminous, and no such star could be detected associated with G2. It was proposed^{7,8} that G2 was a shock front caused by the interaction between a wind from a low-mass star embedded in the cloud and the dense plasma found in the vicinity of the Galactic Centre. It might also have been a protoplanetary disk; or a star ripped apart by tidal forces following previous close encounters with the black hole; or even a merger of two stars.

Armed with hindsight and more data, a few conclusions can now be drawn. The fact that G2 survived the plunge towards the SMBH and emerged intact^{9,10} means that it must be associated with a compact, gravitationally bound object. Its dim infrared emission implies a luminosity less than 30 times that of the Sun. And because it is best detected in the wavelength region between 2 and 5 micrometres, its emission is dominated by that of dust at a temperature of 400 to 600 kelvin. A plausible explanation has been that the G2 cloud contains a young, low-mass star surrounded by a dusty disk that is losing mass, or that it has a wind similar to those found escaping from young low- to intermediate-mass stars near the Sun^{7,10,11}.

Young stars that exhibit excess infrared emission, which is an indicator of extreme youth, have been found in the Galactic Centre before¹². Circumstellar disks irradiated

by ultraviolet radiation, such as those found in the nearby Orion Nebula¹³, or those that are probably associated with young stars in the Galactic Centre, lose mass in a process called photo-ablation. Protoplanets, planetary embryos or planets in dense star clusters can be gravitationally stripped away from their parent stars following encounters with other stars to become free-floating, sub-stellar-mass objects¹⁴.

The occasional encounters of young, low-mass objects with the Milky Way's SMBH may cause them⁶ to brighten dramatically as tidal forces pull them apart. Therefore, events such as the G2 encounter might provide a new way of detecting such otherwise invisible, low-mass objects. Mapelli and Ripamonti model the effects of tidal disruption and photo-ablation on G2 and demonstrate that young planetary embryos with a mass of about 10 to 100 times that of Jupiter, or protoplanets with diameters comparable to the Earth–Sun separation, may become detectable as objects resembling G2. This would occur as their outer layers are stripped away by the SMBH, heat up, and become ionized by the ambient radiation field.

If further analyses show that G2 does not contain a star, the authors' proposal that G2 is a planetary embryo bound together by its relatively weak self-gravity could be a plausible explanation of this mysterious object. Continued monitoring of the Galactic Centre by facilities such as the Atacama Large Millimeter/submillimeter Array in Chile and the Jansky Very Large Array in New Mexico, could complement infrared observations and tell us about the formation¹⁵ of low-mass stars, and even planetary systems, in extreme environments such as our Galaxy's centre. ■

John Bally is in the Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, Colorado 80389, USA.

e-mail: john.bally@colorado.edu

- Gillessen, S. *et al.* *Nature* **481**, 51–54 (2012).
- Mapelli, M. & Ripamonti, E. *Astrophys. J.* **806**, 197 (2015).
- Lu, J. R. *et al.* *Astrophys. J.* **764**, 155 (2013).
- Liu, H. B. *et al.* *Astrophys. J.* **756**, 195 (2012).
- Morris, M. *Nature* **481**, 32–33 (2012).
- Schartmann, M. *et al.* *Astrophys. J.* **755**, 155 (2012).
- Scoville, N. & Burkert, A. *Astrophys. J.* **768**, 108 (2013).
- Ballone, A. *et al.* *Astrophys. J.* **776**, 13 (2013).
- Zajacek, M., Eckart, A., Peissker, F., Karssen, G. D. & Karas, V. Preprint at <http://arxiv.org/abs/1507.00237> (2015).
- Valencia-S., M. *et al.* *Astrophys. J.* **800**, 125 (2015).
- Murray-Clay, R. A. & Loeb, A. *Nature Commun.* **3**, 1049 (2012).
- Eckart, A. *et al.* *Astron. Astrophys.* **551**, A18 (2013).
- Bally, J., O'Dell, C. R. & McCaughrean, M. J. *Astron. J.* **119**, 2919 (2000).
- Wang, L., Kouwenhoven, M. B. N., Zheng, X., Church, R. P. & Davies, M. B. *Mon. Not. R. Astron. Soc.* **449**, 3543 (2015).
- Yusef-Zadeh, F. *et al.* *Astrophys. J.* **801**, L26 (2015).

Structural integration in hypoxia-inducible factors

Dalei Wu¹, Nalini Potluri¹, Jingping Lu¹, Youngchang Kim² & Fraydoon Rastinejad¹

The hypoxia-inducible factors (HIFs) coordinate cellular adaptations to low oxygen stress by regulating transcriptional programs in erythropoiesis, angiogenesis and metabolism. These programs promote the growth and progression of many tumours, making HIFs attractive anticancer targets. Transcriptionally active HIFs consist of HIF- α and ARNT (also called HIF-1 β) subunits. Here we describe crystal structures for each of mouse HIF-2 α -ARNT and HIF-1 α -ARNT heterodimers in states that include bound small molecules and their hypoxia response element. A highly integrated quaternary architecture is shared by HIF-2 α -ARNT and HIF-1 α -ARNT, wherein ARNT spirals around the outside of each HIF- α subunit. Five distinct pockets are observed that permit small-molecule binding, including PAS domain encapsulated sites and an interfacial cavity formed through subunit heterodimerization. The DNA-reading head rotates, extends and cooperates with a distal PAS domain to bind hypoxia response elements. HIF- α mutations linked to human cancers map to sensitive sites that establish DNA binding and the stability of PAS domains and pockets.

Three HIF- α proteins (HIF-1 α , HIF-2 α and HIF-3 α) mediate the adaptive transcriptional response to hypoxia in both normal and tumour cells^{1,2}. To function as productive transcription factors, HIF- α proteins heterodimerize with the aryl hydrocarbon receptor nuclear translocator (ARNT (HIF-1 β))^{3–5}. The HIF- α and ARNT proteins belong to the mammalian basic helix–loop–helix–PER-ARNT-SIM (bHLH-PAS) family^{6,7}, which includes the aryl hydrocarbon receptor (AHR), neuronal PAS proteins (NPAS1, NPAS2, NPAS3, NPAS4), single-minded proteins (SIM1 and SIM2), and the circadian clock components BMAL1 and CLOCK (see Extended Data Fig. 1a). By shaping transcriptional responses to environmental and cellular signals, these factors drive a variety of physiological and disease processes in humans⁸.

The HIF- α proteins function as the oxygen-sensitive regulatory subunits of HIF- α -ARNT heterodimers¹. HIF-1 α and HIF-2 α are critical for the oxygen stress response⁹, while the role of HIF-3 α remains less understood¹⁰. Under normoxia, two proline residues in the oxygen-dependent degradation domain of HIF-1 α and HIF-2 α are hydroxylated by prolyl hydroxylase domain (PHD) enzymes¹¹. PHD activity requires both oxygen availability and the TCA cycle intermediate α -ketoglutarate. The proline modifications target the HIF- α proteins to the ubiquitin–proteasome (26S) pathway for degradation¹². An asparagine residue in their carboxy-terminal segment is also hydroxylated¹³, blocking HIF-1 α interactions with the coactivator p300 (ref. 14).

Under hypoxia, HIF-1 α and HIF-2 α subunits escape degradation and heterodimerize with ARNT to regulate gene programs^{11,15}. These programs increase the oxygen supply and nutrient delivery to cells via angiogenesis and erythropoiesis. The HIFs additionally upregulate the expression of glucose transporters and glycolytic enzymes, to supply the metabolic needs of growing cells in low oxygen conditions. These pathways are particularly important for the growth of many solid tumours. HIF activity is also critical for the progression of kidney-disease-related anaemia, cardiac ischaemia, fatty liver and diabetes^{16–18}. The modulation of HIF pathway by small molecules is recognized as a promising approach in cancer therapeutics^{19,20}.

The PHDs that regulate HIF- α proteins are actively pursued as anticancer targets²¹; however, the multi-domain structures of HIF- α -ARNT complexes have not been visualized, hindering understanding of their drug-binding capabilities.

As members of the bHLH-PAS family⁶, the HIF- α and ARNT proteins have an amino-terminal bHLH segment for DNA binding, tandemly positioned PAS domains (PAS-A and PAS-B) for dimerization, and transactivation domains (TADs)⁷ (Fig. 1a). For our studies, we used their contiguous bHLH-PAS-A-PAS-B segments, which contain all the conserved domains of bHLH-PAS proteins (Extended Data Fig. 1b). Our current understanding of quaternary structure in this family is based only on the CLOCK–BMAL1 structure²², which also used their bHLH-PAS-A-PAS-B segments. The CLOCK–BMAL1 complex was studied in its DNA-free state, and no ligands were available to assess potential pockets. Here, we present a gathering of five distinct crystal structures for mouse HIF-1 α and HIF-2 α heterodimers with ARNT, including the HIF-2 α -ARNT complex alone and bound to two different small molecules, as well as the structures of both HIF-2 α -ARNT and HIF-1 α -ARNT in their DNA-bound states.

Quaternary architecture of HIF-2 α -ARNT heterodimer

We obtained well-diffracting crystals of the HIF-2 α -ARNT heterodimer in the *apo* form, and solved its structure at 2.4 Å resolution (Extended Data Table 1). There are two closely related complexes in the asymmetric unit, with a root mean squared deviation (r.m.s.d.) of less than 1.0 Å between all the main-chain atoms. Figure 1b shows the overall organization of the HIF-2 α -ARNT heterodimer. Each domain from both subunits forms domain–domain interactions to establish a well-integrated quaternary architecture. The three domains of ARNT rotate and twist around the outer surface of the HIF-2 α protein. This form of interaction involves nonequivalent contacts between the two subunits. The asymmetric arrangement of HIF-2 α and ARNT subunits is remarkable, given that each partner employs the same three conserved domains (Fig. 1b and Extended Data Fig. 2).

¹Metabolic Disease Program, Sanford Burnham Prebys Medical Discovery Institute, Orlando, Florida 32827, USA. ²Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA.

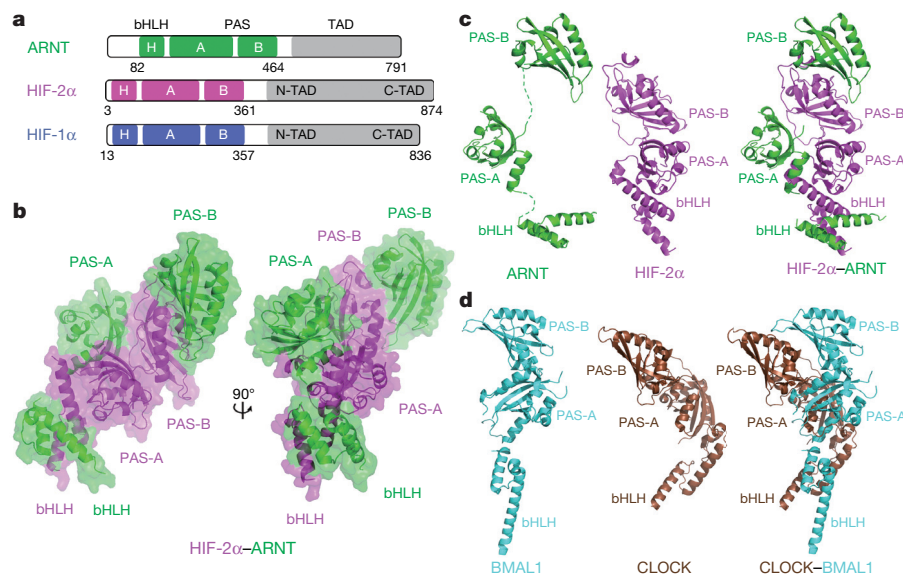


Figure 1 | Overall architectural features of the HIF-2α-ARNT heterodimer and its comparison with CLOCK-BMAL1. **a**, Schematic representation showing the domain arrangements of ARNT (green), HIF-2α (magenta) and HIF-1α (blue) proteins. **b**, Overall structure of the HIF-2α-ARNT heterodimer

Despite their overall asymmetry, the HIF-2α and ARNT subunits bring their respective bHLH segments into convergence to form a single DNA-reading head (Fig. 1a, b). The two bHLH domains become linked in a pseudo-symmetric arrangement, ensuring their ability to read consensus hypoxia response element (HRE) sequences that are also pseudo-symmetric²³. With the DNA-reading head at one end of the complex, the opposite end joins the two PAS-B domains in an asymmetric fashion (Fig. 1b). The central core of the heterodimer is established by the interconnected PAS-A domains from both subunits. PAS domains may function as the signal recognition modules in the bHLH-PAS family^{24,25}.

We separated the two subunits of HIF-2α-ARNT for closer individual inspection, to highlight their important differences (Fig. 1c). HIF-2α has its three domains forming mutual contacts to form a contiguous surface. But the three ARNT domains have no intra-molecular contacts, and are instead rotated and displaced apart (Fig. 1c). The lack of physical constraints among the ARNT domains could confer adaptability for heterodimerization with many other bHLH-PAS members (Extended Data Fig. 1a). All three of the ARNT domains become highly constrained by forming intermolecular contacts within the HIF-2α-ARNT heterodimer (Fig. 1b, c).

Comparison to the CLOCK-BMAL1 architecture

ARNT and BMAL1 are each common dimerization partners for the other mammalian bHLH-PAS members⁶. BMAL1 forms heterodimers with CLOCK and NPAS2. ARNT has a much larger repertoire of partners that includes all three HIF-α proteins, NPAS1, NPAS3, NPAS4, AHR, AHRR, SIM1 and SIM2 (Extended Data Fig. 1a). Next, we asked whether the HIF-2α-ARNT heterodimer shares the same architecture as CLOCK-BMAL1.

We compared the HIF-2α-ARNT and CLOCK-BMAL1 heterodimers and their individual proteins (Fig. 1c, d). First we find that while ARNT has its three domains separated, BMAL1 has connected PAS-A and PAS-B domain surfaces. Then comparing HIF-2α with CLOCK, we find different domain-domain arrangements in each protein (Fig. 1c, d). The differences among all four proteins are striking, since they share conserved and similarly embedded bHLH, PAS-A and PAS-B domains (Extended Data Fig. 1b). We next compared HIF-2α-ARNT and CLOCK-BMAL1 at the level of heterodimers, finding notable differences in their dimerization modes (Fig. 1c, d and

Extended Data Fig. 3). Given that each of BMAL1 and ARNT heterodimerize with non-overlapping subsets of bHLH-PAS proteins, these two distinct arrangements may represent how the two subsets of family members are architecturally arranged.

We note that three of the domain-domain interfaces are similarly formed within the HIF-2α-ARNT and CLOCK-BMAL1 heterodimers (see interfaces 1, 2 and 4 in Extended Data Fig. 4a, b). But a previously described interaction mode observed by NMR²⁶ and crystallography²⁷, involving the isolated PAS-B domains of HIF-2α and ARNT, is not seen in our structure (Extended Data Fig. 4a, c). One cannot exclude that the previously observed mode of interaction is relevant during the assembly of the complex or upon recruitment of associated factors. A binding site for coactivator TACC3 on the surface of ARNT PAS-B domain has also been described recently²⁸, which is not fully accessible within our multi-domain structure (Extended Data Fig. 4d).

Similar domain interfaces within two HIF heterodimers

We next examined whether the crystallographic observations of domain-domain interfaces in HIF-2α-ARNT are important for establishing heterodimer stability in cells. Figure 2a shows all six domain-domain interfaces observed in our crystal structure in detail. We probed these interfaces using a mutational strategy directed at the amino-acid side chains that form the interfacial contacts. The stabilities of the heterodimers within HEK293T cells were examined using a co-immunoprecipitation strategy that detects if the two subunits still assemble when the mutations are present (Fig. 2b).

We found that most single and double mutations introduced at the observed domain-domain surfaces produced significantly destabilizing effects on the HIF-2α-ARNT heterodimer (Fig. 2b). Mutations at interface 1, connecting the ARNT bHLH to the HIF-2α bHLH domain, and at interface 2, connecting ARNT PAS-A to the HIF-2α PAS-A, clearly compromised heterodimer stability (Fig. 2b). The domain connection linking the ARNT PAS-A to the HIF-2α PAS-B (interface 3) is a critical part of our structure, but the equivalent interface does not form in the CLOCK-BMAL1 structure²². Consistently, single and double mutations introduced at interface 3 destabilized the HIF-2α-ARNT heterodimer. Further mutations introduced at other domain-domain junctions also reduced heterodimer stability (Fig. 2b). These observations suggest that the domain

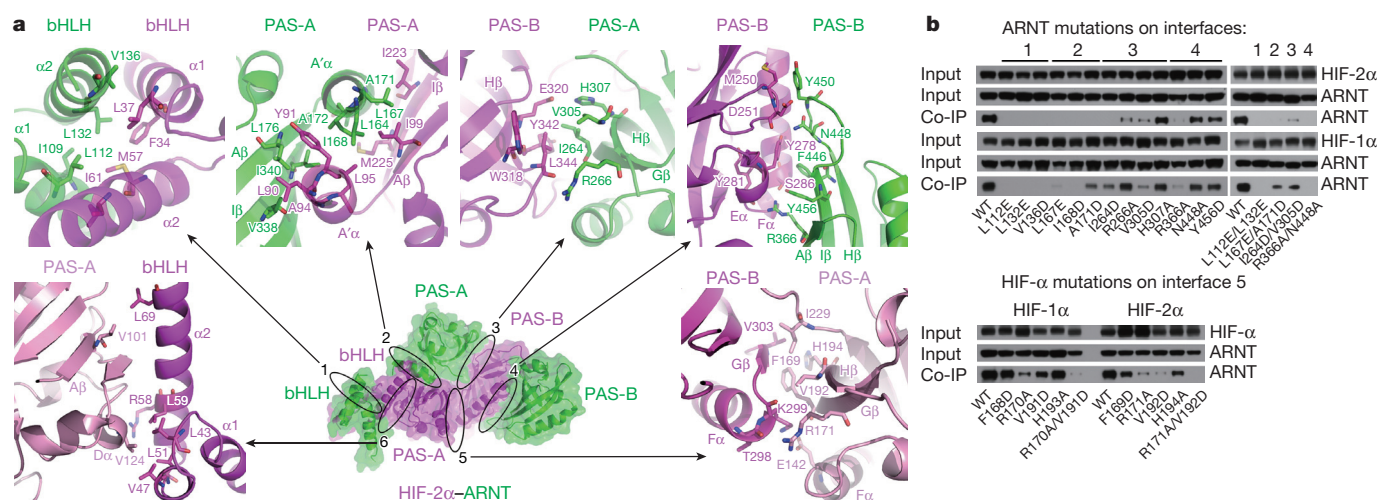


Figure 2 | Domain interfaces of the HIF-2 α -ARNT complex. **a**, Overall structure of HIF-2 α -ARNT with each domain labelled and domain interfaces numbered 1 to 6 (lower centre); and detailed interactions at those interfaces in close-up views (surrounding panels). Interfaces 1–4 are formed between ARNT (green) and HIF-2 α (magenta), while 5 and 6 are intra-molecular

junctions in HIF-2 α . **b**, Co-immunoprecipitation (co-IP) experiments showing the effects of ARNT mutations at interfaces 1–4 on the stability of ARNT heterodimers with HIF-2 α or HIF-1 α . Interface 5 was disrupted by mutations in HIF-1 α or HIF-2 α at their PAS-A domains. Full blot scans are shown in the Supplementary Fig. 1.

interfaces observed in the crystal structure are indeed required for heterodimerization between HIF-2 α and ARNT in cells.

HIF-1 α and HIF-2 α have a much closer level of amino acid homology to each other than they do to other family members (Extended Data Fig. 1b). Our data show that mutations that destabilized HIF-2 α -ARNT also destabilized HIF-1 α -ARNT (Fig. 2b). We infer from these data that the two heterodimers have closely related architectures. Our subsequent structural observations of both HIF-1 α -ARNT and HIF-2 α -ARNT in their DNA-bound states confirm this notion (see below).

Distinct pockets for small-molecule binding

Among the mammalian bHLH-PAS family members, AHR is known to function physiologically as a sensor of small molecules, binding to a diverse set of xenobiotic and endogenous molecules²⁹. Another member, NPAS2, has been reported to be gas responsive and bind haem as a prosthetic group²⁵. It is not known whether other members of this family have natural ligands. PAS domains in unrelated signal transduction protein families are known to act as versatile molecular sensors, to detect physical, chemical and environmental signals^{30,31}.

Although no endogenous ligand has been identified for the HIF- α or ARNT proteins, a number of synthetic small molecules were discovered through screening campaigns^{32–36}. Ligands that reduce the activity of HIF- α proteins are sought as potentially useful therapies in cancers where HIF- α overexpression correlates inversely with patient survival^{9,11}. The possibility that HIF- α proteins contain a pocket was initially motivated by the discovery of a large cavity filled with eight ordered water molecules in the isolated PAS-B domain from HIF-2 α (refs 26, 27, 37).

We directly investigated the binding modes of two previously described molecules (OX3 and acriflavine) within the HIF-2 α -ARNT heterodimer using X-ray crystallography (Fig. 3a–e). The binding site for acriflavine had not been known. However, acriflavine acts through the HIF- α proteins to inhibit tumour vascularization and mobilization of angiogenic cells into peripheral blood³⁸. For OX3, the binding site within the isolated PAS-B domain of HIF-2 α was previously examined³⁴, and we asked whether its binding pocket is shared by acriflavine.

Acriflavine contains a mixture of two structurally related components, proflavine and tryptaflavin³⁸. We began by showing that both components bind to HIF-1 α and HIF-2 α heterodimers with ARNT with similar affinities (Extended Data Fig. 5a, b). We obtained crystals

of HIF-2 α -ARNT with each of OX3 and proflavine, then solved and refined these structures at ~ 2.8 Å resolution (Extended Data Table 1). The electron density maps in Fig. 3b–e allowed us to visualize their binding sites unambiguously.

We observed that OX3 and proflavine use distinct binding pockets within the architecture (Fig. 3a). OX3 becomes fully encapsulated within the PAS-B domain of HIF-2 α (Fig. 3b, c). Therefore, the quaternary architecture of the heterodimer formed no apparent barriers for its access to that site. Proflavine binds to an unexpected site. It intercalates between the HIF-2 α PAS-B domain and the ARNT PAS-A domain (Fig. 3d, e). The creation of pockets between PAS domains had not been anticipated. Moreover, the binding site of proflavine is at the critical convergence point of these two domains (interface 3, Figs 2a and 3d, e).

Amino acid residues R266 and V305 of ARNT form direct contacts with proflavine (Fig. 3e), but are also critical for maintaining the integrity of interface 3 (Fig. 2a). Our mutagenesis studies in Fig. 2b showed that this site is a critical node required for establishing stable HIF- α -ARNT heterodimers. Accordingly, the binding of acriflavine was reported to destabilize HIF- α -ARNT heterodimers in cells³⁸. A previous study showed that acriflavine can also bind to each of the individual HIF-1 α and HIF-2 α subunits with a dissociation constant (K_d) of ~ 50 μ M³⁸. This binding is substantially weaker than the K_d of ~ 40 nM we observed for binding to the HIF-1 α -ARNT and HIF-2 α -ARNT heterodimers (Extended Data Fig. 5b). Nevertheless, it remains possible that acriflavine binding can occur at a second independent site on the HIF- α subunit alone. The PAS-B domain in HIF- α is another critical node for establishing the stability of heterodimers, as it firmly interacts with two PAS domains from ARNT, as well as HIF- α 's own PAS-A domain (Figs 2a and 3a). We show in Extended Data Fig. 5c a plausible mechanism through which ligand binding to the HIF-2 α PAS-B domain can influence heterodimer stability in the HIF-2 α -ARNT complex.

We examined whether the amino acid residues in the PAS-B pockets of HIF-2 α and HIF-1 α are strictly conserved, or sufficiently different to allow small-molecule discrimination (Extended Data Fig. 5d, e). Three OX3-interacting residues differ in these two pockets: A277, S292 and S304 in HIF-2 α are replaced by I275, T290 and T302 (which have bulkier side chains) in HIF-1 α . As previously suggested^{32,34}, these differences cause a smaller cavity to form in HIF-1 α , explaining why OX3 binds to the HIF-2 α PAS-B domain selectively. Another small molecule, THS-044, also shows selective binding to the HIF-2 α PAS-B

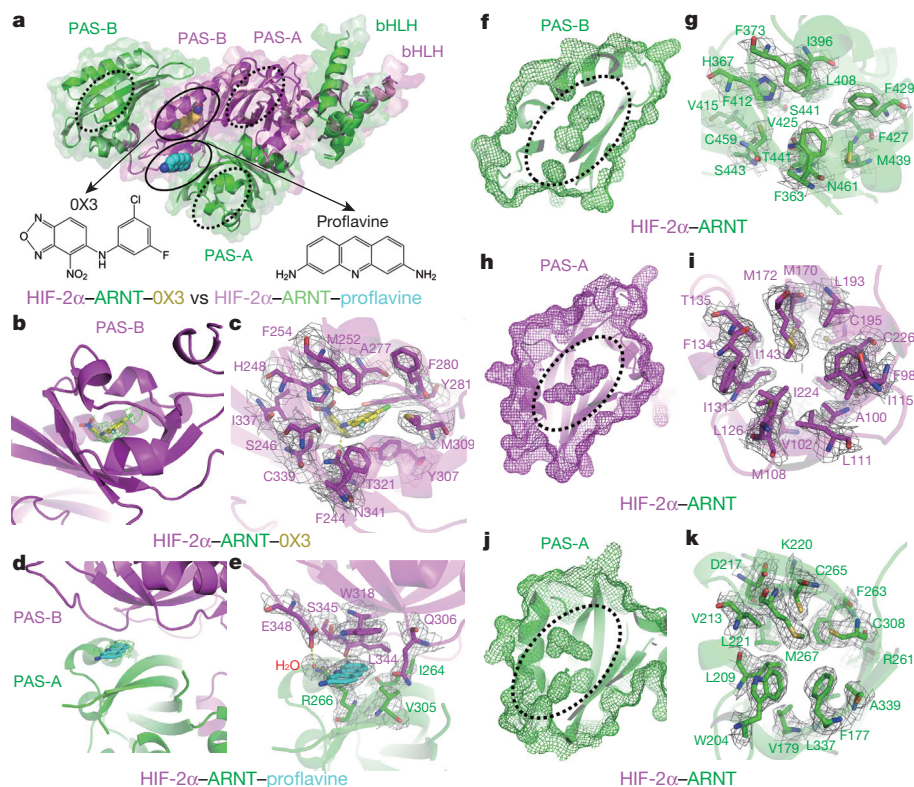


Figure 3 | Ligand binding pockets in the HIF-2α-ARNT heterodimer. **a**, Superimposition of OX3 and proflavine binding positions in the overall HIF-2α-ARNT architecture. The solid ovals indicate observed locations for these compounds, and the dotted ovals show the positions of empty pockets. **b–e**, The locations and interaction of OX3 (**b**, **c**) and proflavine (**d**, **e**), with green meshes showing $F_o - F_c$ SA omit map contoured at 2.4σ (**b**, **d**) and grey meshes

showing $2F_o - F_c$ electron density at 1.0σ (**c**, **e**). **f–k**, Three empty pockets visualized inside the ARNT PAS-B (**f**, **g**), HIF-2α PAS-A (**h**, **i**) and ARNT PAS-A (**j**, **k**) domains. The meshes inside the dotted ovals depict the accessible cavities within each domain (**f**, **h**, **j**), and the surrounding residues that line these pockets are covered with $2F_o - F_c$ map at 1.0σ (**g**, **i**, **k**).

domain²⁷. However, the observed proflavine pocket uses a nearly identical set of amino acids within HIF-2α-ARNT and HIF-1α-ARNT heterodimers (Extended Data Fig. 5f, g). Consistent with this observation, acriflavine was found to interact with both HIF forms³⁸, which we confirmed biochemically (Extended Data Fig. 5b).

We also looked elsewhere for accessible small-molecule binding pockets in the HIF-2α-ARNT architecture. As previously reported³⁶, the PAS-B domain of ARNT contains another available cavity (Fig. 3f, g). The small molecule KG-548 has been shown to bind there, inhibiting its interactions with the coactivator TACC3³⁶. Only some residues are conserved between the ARNT and HIF-2α PAS-B pockets. Consistent with this observation, KG-548 binds only to ARNT, but not to HIF-2α (ref. 36). The PAS-B pocket of ARNT represents a third distinct pocket in the heterodimer for small-molecules binding.

In addition, we also found two other accessible cavities within the HIF-2α-ARNT complex. These are observed in the PAS-A domains of each of HIF-2α and ARNT proteins (Fig. 3h–k). It was reported previously that the NPAS2 protein binds haem through both its PAS-A and PAS-B domains²⁵, and some proteins outside the bHLH-PAS family use their PAS-A domains for ligand binding^{30,31}. Importantly, only about 60% of residues embracing the HIF-2α PAS-A pocket are identical with HIF-1α. Therefore, this pocket may offer selective binding to ligands between the two HIF-α proteins, just as with their PAS-B pockets. The volumes of all the PAS pockets of HIF-2α and ARNT are shown in Extended Data Fig. 5h. These pockets can potentially bind endogenous molecules or synthetic compounds. The discovery of five distinct pockets in these complexes expands the possibilities for targeting HIF protein activities with drug-like molecules for therapeutic purposes.

Binding to hypoxia response elements

To understand how HIF-1α-ARNT and HIF-2α-ARNT complexes recognize HREs, we prepared segments consisting of the bHLH-PAS-A-PAS-B domains from HIF-1α, HIF-2α and ARNT, together with synthetic DNA duplex containing the core sequence of HRE (5'-TACGTG-3'). Crystal structures for HIF-1α-ARNT and HIF-2α-ARNT complexes with DNA were solved and refined at 3.9 Å and 3.6 Å resolutions, respectively (Extended Data Table 1).

These structures confirm that the overall architectures of two HIF-α-ARNT heterodimers are indistinguishable (Fig. 4a, b), as our mutagenesis studies suggested (Fig. 2b). Moreover, HIF-1α-ARNT and HIF-2α-ARNT engage their response elements in an identical fashion, using specific amino acids located on their bHLH domains. These domains insert α-helices into the major groove faces of the TACGTG sequences in a pseudo-symmetric fashion, allowing base-pair-specific hydrogen bonds and van der Waals interactions to form with residues from the bHLH helices (Extended Data Fig. 6a, b). The ARNT residues forming DNA contacts (R102, E98 and H94) were identically positioned in BMAL1 to recognize the GTG half site, as revealed by the crystal structure of isolated CLOCK-BMAL1 bHLH domain with E-box (5'-CACGTG-3') DNA³⁹. We further verified the importance of these residues in HIF-2α-ARNT by introducing mutations and detecting reduced DNA binding as a result (Extended Data Fig. 6c).

A comparison of the DNA-bound and DNA-free forms of HIF-2α-ARNT shows important structural adaptations. As seen in Fig. 4c, the α1 helices in both HIF-2α and ARNT bHLH domains extend in size to bind to DNA. The α2 helices, which link the bHLH domains to the remainder of the quaternary architecture, undergo rotations to further facilitate DNA binding. The extension and rotation of the

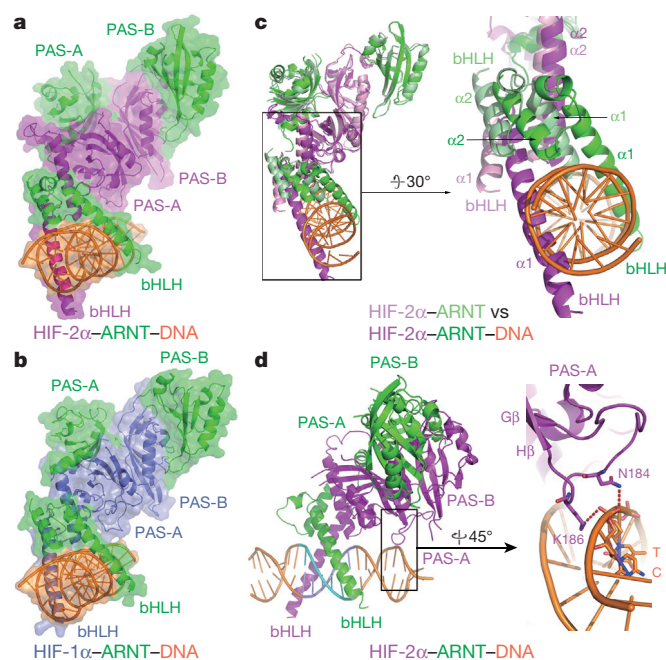


Figure 4 | DNA-bound HIF- α -ARNT structures. **a, b**, Overall view of the HIF-2 α -ARNT-DNA (**a**) and HIF-1 α -ARNT-DNA (**b**) complexes. **c**, Comparison of the DNA-free and DNA-bound HIF-2 α -ARNT heterodimers (left); enlarged figure shows the dramatic conformational changes at the bHLH domains that accompany DNA binding (right). **d**, The interactions between HIF-2 α PAS-A domain and DNA (HRE site in blue) are shown in an overview (left) and in a close-up fashion (right).

DNA-reading head enhances the heterodimer's ability to engage the DNA response element optimally.

In both heterodimers, we also found that the PAS-A domain of HIF- α cooperates with the bHLH domains to establish DNA binding (Fig. 4d). Mutations introduced at the DNA-binding surface of the HIF-2 α PAS-A domain impacted DNA affinity (Extended Data Fig. 6c, d). The PAS-A domain contacts DNA at a distance six base pairs away from the hexameric core element (Fig. 4d and Extended Data Fig. 6b). This cooperation between the PAS-A and bHLH domains markedly extends the heterodimers' footprints on DNA. PAS domains had not been known to be capable of forming DNA interactions. Specifically positioned amino acids within the PAS-A domains of ARNT's partners may modulate the DNA-binding affinities of other bHLH-PAS heterodimers.

Mapping cancer mutations and phosphorylation sites

There is strong evidence for altered HIF functions in many human cancers¹⁵. Germline mutations in the von Hippel-Lindau (*VHL*) tumour suppressor gene, encoding a HIF- α E3 ubiquitin protein ligase, cause a dominantly inherited familial cancer syndrome⁴⁰. The resulting mutational inactivation of VHL protein leads to the accumulation of HIF-1 α and HIF-2 α proteins^{41,42}. VHL disease is characterized by susceptibility to retinal and central nervous system haemangioblastomas, clear cell renal cell carcinomas (ccRCC), pancreatic islet cell tumours and renal, pancreatic and epididymal cysts⁴³.

The *HIF1A* gene is highly overexpressed in a number of cancers including colon, breast, gastric, lung, skin, ovarian, pancreatic, prostate and renal cancers as a result of intratumoral hypoxia¹⁹. In the case of renal cancer, recent genetic and functional studies also implicated HIF-1 α as a tumour suppressor gene⁴⁴. Loss of HIF-1 α expression was found to be associated with worse survival rate in patients with ccRCC⁴⁵. Changes in expression levels of HIF- α proteins present in a wide spectrum of cancers occur through a multitude of

mechanisms, including hypoxia, loss-of-function of tumour suppressor genes, activating mutations in oncogenes, and activation of kinase signalling pathways¹⁵.

Missense mutations within each of HIF-1 α and HIF-2 α proteins have been isolated from different cancer biopsies^{46,47}. Using the COSMIC (Catalogue of Somatic Mutations in Cancer) database⁴⁸, we mapped 19 such mutations that locate to the bHLH-PAS-A-PAS-B segments of these proteins (Extended Data Fig. 7a). These mutations are linked to cancers including stomach adenocarcinomas, endometrial carcinomas, brain gliomas, lung adenocarcinomas, hepatocellular carcinomas and skin melanomas.

Extended Data Figure 7 shows the positions of these mutations within both heterodimers. One group of mutations maps directly to sites that establish DNA binding. A second group maps to the domain interfaces, while the rest locate to the pockets in PAS domains. Notably, some mutations at PAS-B domains specifically alter the interior pockets where we observed 0X3 binding. These include residues H248 and A277 inside the HIF-2 α pocket (Extended Data Fig. 7b), which contact 0X3 (Fig. 3c and Extended Data Fig. 5d). M250I alters the equivalent PAS-B pocket within HIF-1 α (Extended Data Figs 5e and 7c). The HIF-1 α mutation V116E identified from ccRCC patients has been shown to impair HIF-1 α activity⁴⁷. As shown in Extended Data Fig. 7c, V116 is positioned on an α -helix (C α) in the PAS-A domain, where its hydrophobic side chain is required to stabilize this domain. Phosphorylation can also regulate HIF- α activity on some target genes⁹. HIF-1 α and HIF-2 α are phosphorylated in their PAS-B domains^{49,50}. We examined these sites within the heterodimeric complexes (Extended Data Fig. 8).

Concluding remarks

Hypoxia in solid tumours contributes to aggressive phenotypes and accounts for the failure of many current therapies⁹. The HIFs are versatile transcription factors responsible for progression of hypoxia-induced tumours¹¹. HIF-regulated programs drive erythropoiesis, angiogenesis and the metabolic switch to glucose utilization for fuel. The HIFs further drive drug efflux, cell proliferation and anti-apoptosis programs¹⁵. The coordination of all these programs helps to promote tumour growth, drug resistance and ultimately leads to lower patient survivability¹⁹. As many classic chemotherapy agents fail to impact hypoxic cancers, novel strategies are needed to target tumour growth under hypoxic conditions²⁰. Targeting of the HIF-1 α and HIF-2 α proteins with a broad variety of therapeutic small-molecule modulators should be possible given the five distinct pockets visualized here. As the HIF pathways are also critical for driving a number of other diseases in humans, including anaemia and cardiac ischaemia, the discovery of drug-binding pockets opens the possibility for therapeutic approaches directed at the HIF proteins themselves, in place of the HIF-modifying PHD enzymes.

Other ligand-responsive mammalian bHLH-PAS family members include AHR²⁴ and NPAS2²⁵. These factors bind to environmental and cellular ligands. The HIF- α proteins may also recognize endogenous ligands that provide cues about the environmental or metabolic status of cells. The observation that several cancer-related missense mutations map to these pockets suggests that endogenous ligand(s) are likely to exist. Identifying these natural ligands and their effects on the heterodimers would further propel our understanding of HIF functions and drive the discovery of small-molecule modulators. It is intriguing that some pockets form at critical nodes that establish heterodimer stability (Extended Data Fig. 9). Protein stability is already a well-established mechanism by which the HIF- α proteins are regulated physiologically⁹.

Finally, we point to the differences between the HIF- α -ARNT and CLOCK-BMAL1 architectures as being indicative that a single mode of heterodimerization is not shared by all mammalian bHLH-PAS members. HIF- α -ARNT and CLOCK-BMAL1 heterodimers employ uniquely arranged domain-domain junctions to establish

their stabilities. Their integrated subunits and domain arrangements provide the pathways through which signals may be communicated allosterically across domains. These signals can include endogenous ligands, disease mutations and post-translational modifications. Future studies that explore the allosteric pathways in these heterodimers can also advance the discoveries of therapeutic molecules.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 March; accepted 7 July 2015.

Published online 5 August 2015.

1. Semenza, G. L. Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 (1999).
2. Semenza, G. L. Hypoxia-inducible factors in physiology and medicine. *Cell* **148**, 399–408 (2012).
3. Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl Acad. Sci. USA* **92**, 5510–5514 (1995).
4. Jiang, B. H., Rue, E., Wang, G. L., Roe, R. & Semenza, G. L. Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J. Biol. Chem.* **271**, 17771–17778 (1996).
5. Peng, J., Zhang, L., Rysdale, L. & Fong, G. H. The transcription factor EPAS-1/hypoxia-inducible factor 2alpha plays an important role in vascular remodeling. *Proc. Natl Acad. Sci. USA* **97**, 8386–8391 (2000).
6. Bersten, D. C., Sullivan, A. E., Peet, D. J. & Whitelaw, M. L. bHLH-PAS proteins in cancer. *Nature Rev. Cancer* **13**, 827–841 (2013).
7. Kewley, R. J., Whitelaw, M. L. & Chapman-Smith, A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell Biol.* **36**, 189–204 (2004).
8. McIntosh, B. E., Hogenesch, J. B. & Bradfield, C. A. Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annu. Rev. Physiol.* **72**, 625–645 (2010).
9. Keith, B., Johnson, R. S. & Simon, M. C. HIF1 α and HIF2 α : sibling rivalry in hypoxic tumour growth and progression. *Nature Rev. Cancer* **12**, 9–22 (2012).
10. Heikkilä, M., Pasanen, A., Kivirikko, K. I. & Myllyharju, J. Roles of the human hypoxia-inducible factor (HIF)-3alpha variants in the hypoxia response. *Cell. Mol. Life Sci.* **68**, 3885–3901 (2011).
11. Harris, A. L. Hypoxia—a key regulatory factor in tumour growth. *Nature Rev. Cancer* **2**, 38–47 (2002).
12. Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337–1340 (2001).
13. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J. & Whitelaw, M. L. Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science* **295**, 858–861 (2002).
14. Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Structural basis for Hif-1 α /CBP recognition in the cellular hypoxic response. *Proc. Natl Acad. Sci. USA* **99**, 5271–5276 (2002).
15. Semenza, G. L. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J. Clin. Invest.* **123**, 3664–3671 (2013).
16. Semenza, G. L. Hypoxia-inducible factor 1 and cardiovascular disease. *Annu. Rev. Physiol.* **76**, 39–56 (2014).
17. Girgis, C. M., Cheng, K., Scott, C. H. & Gunton, J. E. Novel links between HIFs, type 2 diabetes, and metabolic syndrome. *Trends Endocrinol. Metab.* **23**, 372–380 (2012).
18. Eltzschig, H. K., Bratton, D. L. & Colgan, S. P. Targeting hypoxia signalling for the treatment of ischaemic and inflammatory diseases. *Nature Rev. Drug Discov.* **13**, 852–869 (2014).
19. Semenza, G. L. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. *Oncogene* **29**, 625–634 (2010).
20. Semenza, G. L. Hypoxia-inducible factors: mediators of cancer progression and targets for cancer therapy. *Trends Pharmacol. Sci.* **33**, 207–214 (2012).
21. Hewitson, K. S. & Schofield, C. J. The HIF pathway as a therapeutic target. *Drug Discov. Today* **9**, 704–711 (2004).
22. Huang, N. et al. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189–194 (2012).
23. Wenger, R. H., Stiehl, D. P. & Camenisch, G. Integration of oxygen signaling at the consensus HRE. *Sci. STKE* **2005**, re12 (2005).
24. Murray, I. A., Patterson, A. D. & Perdew, G. H. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature Rev. Cancer* **14**, 801–814 (2014).
25. Dioum, E. M. et al. NPAS2: a gas-responsive transcription factor. *Science* **298**, 2385–2387 (2002).
26. Erbel, P. J., Card, P. B., Karakuzu, O., Bruick, R. K. & Gardner, K. H. Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *Proc. Natl Acad. Sci. USA* **100**, 15504–15509 (2003).
27. Scheuermann, T. H. et al. Artificial ligand binding within the HIF2 α PAS-B domain of the HIF2 transcription factor. *Proc. Natl Acad. Sci. USA* **106**, 450–455 (2009).
28. Guo, Y., Scheuermann, T. H., Partch, C. L., Tomchick, D. R. & Gardner, K. H. Coiled-coil coactivators play a structural role mediating interactions in hypoxia inducible factor heterodimerization. *J. Biol. Chem.* **290**, 7707–7721 (2015).
29. Denison, M. S., Soshilov, A. A., He, G., DeGroot, D. E. & Zhao, B. Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.* **124**, 1–22 (2011).
30. Möglich, A., Ayers, R. A. & Moffat, K. Structure and signaling mechanism of Per-Arnt-Sim domains. *Structure* **17**, 1282–1294 (2009).
31. Henry, J. T. & Crosson, S. Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annu. Rev. Microbiol.* **65**, 261–286 (2011).
32. Cardoso, R. et al. Identification of Cys255 in HIF-1 α as a novel site for development of covalent inhibitors of HIF-1 α /ARNT PasB domain protein-protein interaction. *Protein Sci.* **21**, 1885–1896 (2012).
33. Rogers, J. L. et al. Development of inhibitors of the PAS-B domain of the HIF-2 α transcription factor. *J. Med. Chem.* **56**, 1739–1747 (2013).
34. Scheuermann, T. H. et al. Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nature Chem. Biol.* **9**, 271–276 (2013).
35. Miranda, E. et al. A cyclic peptide inhibitor of HIF-1 heterodimerization that inhibits hypoxia signaling in cancer cells. *J. Am. Chem. Soc.* **135**, 10418–10425 (2013).
36. Guo, Y. et al. Regulating the ARNT/TACC3 axis: multiple approaches to manipulating protein/protein interactions with small molecules. *ACS Chem. Biol.* **8**, 626–635 (2013).
37. Key, J., Scheuermann, T. H., Anderson, P. C., Daggett, V. & Gardner, K. H. Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. *J. Am. Chem. Soc.* **131**, 17647–17654 (2009).
38. Lee, K. et al. Acriflavine inhibits HIF-1 dimerization, tumor growth, and vascularization. *Proc. Natl Acad. Sci. USA* **106**, 17910–17915 (2009).
39. Wang, Z., Wu, Y., Li, L. & Su, X. D. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell Res.* **23**, 213–224 (2013).
40. Latif, F. et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317–1320 (1993).
41. Maxwell, P. H. et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
42. Li, L. et al. Hypoxia-inducible factor linked to differential kidney cancer risk seen with type 2A and type 2B VHL mutations. *Mol. Cell. Biol.* **27**, 5381–5392 (2007).
43. Kaelin, W. G. Jr. Molecular basis of the VHL hereditary cancer syndrome. *Nature Rev. Cancer* **2**, 673–682 (2002).
44. Shen, C. et al. Genetic and functional studies implicate HIF1alpha as a 14q kidney cancer suppressor gene. *Cancer Discov.* **1**, 222–235 (2011).
45. Kroeger, N. et al. Deletions of chromosomes 3p and 14q molecularly subclassify clear cell renal cell carcinoma. *Cancer* **119**, 1547–1554 (2013).
46. Ollerenshaw, M., Page, T., Hammonds, J. & Demaine, A. Polymorphisms in the hypoxia inducible factor-1 α gene (HIF1A) are associated with the renal cell carcinoma phenotype. *Cancer Genet. Cytogenet.* **153**, 122–126 (2004).
47. Morris, M. R. et al. Mutation analysis of hypoxia-inducible factors HIF1A and HIF2A in renal cell carcinoma. *Anticancer Res.* **29**, 4337–4343 (2009).
48. Forbes, S. A. et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
49. To, K. K., Sedelnikova, O. A., Samons, M., Bonner, W. M. & Huang, L. E. The phosphorylation status of PAS-B distinguishes HIF-1alpha from HIF-2 α in NBS1 repression. *EMBO J.* **25**, 4784–4794 (2006).
50. Kalousi, A. et al. Casein kinase 1 regulates human hypoxia-inducible factor HIF-1. *J. Cell Sci.* **123**, 2976–2986 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Zhang and M. Wang for isolation and identification of trypaflavin.

Author Contributions D.W. and F.R. conceived the study; D.W. isolated the proteins, carried out crystallizations and conducted biochemical studies; N.P. produced the expression and mutation constructs; Y.K. and J.L. collected synchrotron diffraction data; D.W., Y.K. and F.R. analysed the data; F.R. and D.W. wrote the manuscript.

Author Information Coordinates and structure factors have been deposited in Protein Data Bank under accession numbers 4ZP4 (HIF-2 α -ARNT apo), 4ZQD (HIF-2 α -ARNT-OX3), 4ZPH (HIF-2 α -ARNT-Proflavine), 4ZPK (HIF-2 α -ARNT-DNA) and 4ZPR (HIF-1 α -ARNT-DNA). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.R. (frastinejad@sbdpcdiscovery.org).

METHODS

No statistical methods were used to predetermine sample size.

Plasmid construction and site-directed mutagenesis. For the protein overexpression in *Escherichia coli*, mouse ARNT (GenBank accession AAH12870.1, residues 82–464) was cloned into the vector pMKH as previously described⁵¹. Meanwhile, mouse HIF-2 α (GenBank AAH57870.1, residues 3–361) and HIF-1 α (GenBank AAH26139.1, residues 13–357) were cloned into the vector pSJ2, respectively. For the co-immunoprecipitation, full-length ARNT was cloned into the pCMV-Tag1 vector (C-terminal Flag-tagged), while full-length HIF-2 α and HIF-1 α were cloned into the pCMV-Tag4 vector (C-terminal Myc-tagged), respectively. Site-directed mutagenesis was performed as previously⁵¹, and confirmed by DNA sequencing.

Protein expression and purification. The recombinant plasmids pSJ2-HIF-2 α and HIF-1 α were respectively co-transformed along with pMKH-ARNT into BL21 competent cells (Agilent Technologies). Following the IPTG-induced overnight expression at 16 °C, cell pellets were lysed by sonication, and supernatant samples after centrifugation were applied onto the packed His·Bind resin (Novagen). The bound proteins were further purified by an SP Sepharose column (GE Healthcare), and the eluted fractions were loaded on a Superdex 200 pg gel filtration column. To prepare protein–DNA complexes, 21mer double-strand DNA (forward: 5'-GGCTGCGTACGTGCGGGTCGT-3' and reverse: 5'-CACGACCCGCACGTA CGCAGC-3') was mixed with the heterodimeric proteins at a molar ratio of 1.2:1. Then the mixture was loaded again on the gel filtration column in the running buffer containing 20 mM Tris pH 8.0 and 150 mM NaCl. Peak fractions of the protein–DNA complex were pooled and supplemented with 10 mM DTT.

Crystallization and X-ray data collection. Crystallization of the HIF-2 α -ARNT complex was carried out using the sitting-drop vapour diffusion method at 16 °C, by mixing equal volume of protein (4 mg ml⁻¹) and reservoir solution containing 2% tacsimate pH 7.0, 6% PEG3350. Before being flash frozen in liquid nitrogen, crystals were soaked in reservoir plus 30% PEG400 as the cryoprotectant. HIF-2 α -ARNT-0X3 (antagonist 2, Sigma SML0883) complex was co-crystallized in a similar condition with 100 μ M compound added to the protein solution. HIF-2 α -ARNT-proflavine crystals were obtained by soaking *apo* crystals in the drops containing 100 μ M acriflavine (Sigma A8126) overnight. HIF-2 α -ARNT-DNA crystals were grown in sitting drops formed by equal volume of complex (4 mg ml⁻¹) and reservoir consisting of 4% tacsimate pH 6.0, 10% PEG3350, and then transferred stepwise to cryoprotectant containing up to 30% PEG 400 before being flash frozen. HIF-1 α -ARNT-DNA crystals were grown also at 16 °C in hanging drops, with the reservoir containing 100 mM MES pH 6.0, 100 mM CaAc₂ and 15% PEG 400. Before being flash frozen, these crystals were air dehydrated by serially transferring to 40 μ l drops of reservoir with increasing concentrations of PEG 400 (up to 40%) for about 10 min at each step. Diffraction data were collected at the Argonne National Laboratory SBC-CAT 19ID beamline at 100 K, and processed using the HKL3000 program⁵².

Structure determination and refinement. The structure of HIF-2 α -ARNT complex was solved by molecular replacement with the program Phaser⁵³, using modified single domains from coordinates 4F3L²², 3F1P²⁷ and 4M4X⁵¹ as the search models. Further manual model building was facilitated by using Coot⁵⁴, combined with the structure refinement using phenix.refine⁵⁵ and the PDB_REDO web server⁵⁶. The rest of the structures were then solved by molecular replacement using HIF-2 α -ARNT complex as the starting model, and further refined in the similar way. The diffraction data and final refinement statistics are summarized in Extended Data Table 1. The Ramachandran statistics, calculated by Molprobity⁵⁷, are 96%/0, 92%/0, 93%/0, 88%/0.35%, 82%/0.41% (favoured/outliers) for the HIF-2 α -ARNT *apo*, HIF-2 α -ARNT-0X3, HIF-2 α -ARNT-proflavine, HIF-2 α -ARNT-DNA and HIF-1 α -ARNT-DNA complexes, respectively. All the structural figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

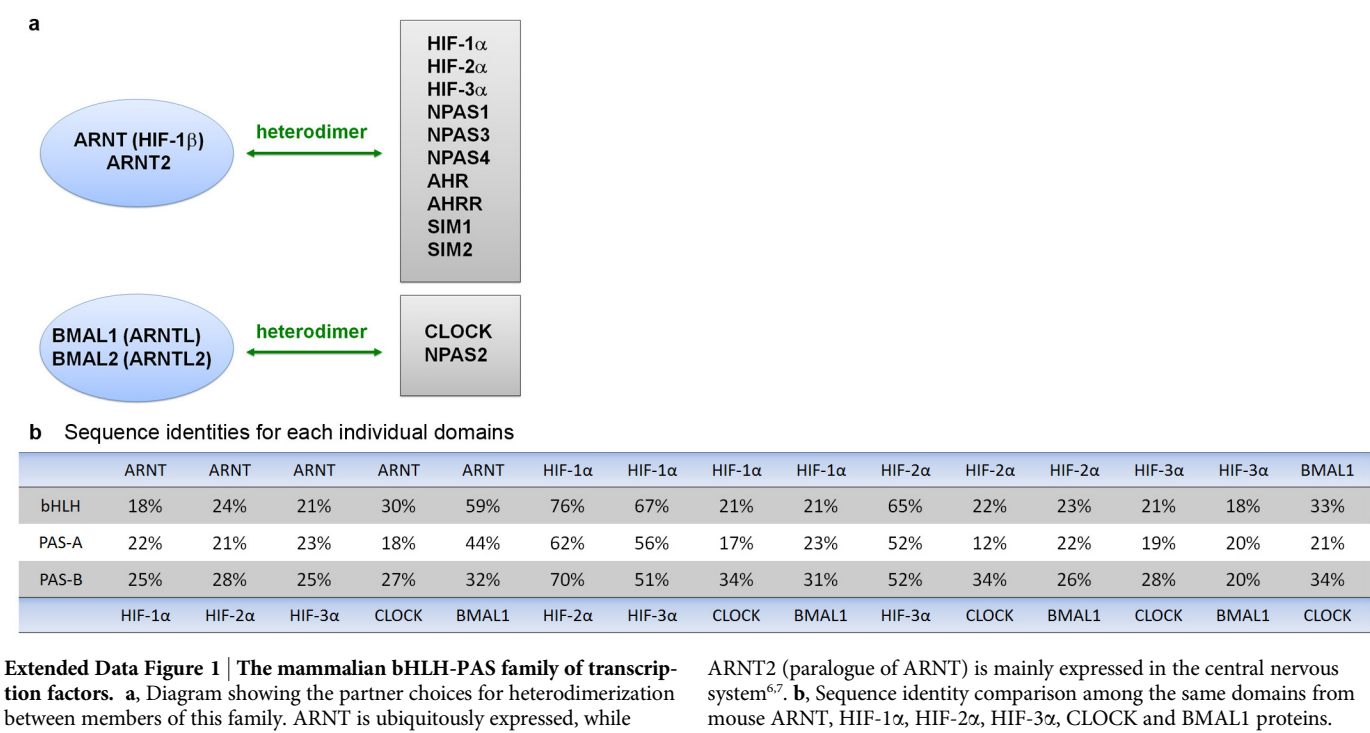
Co-immunoprecipitation. Similarly to our previous work⁵¹, HEK293T cells (ATCC CRL-3216, not authenticated nor tested for mycoplasma contamination)

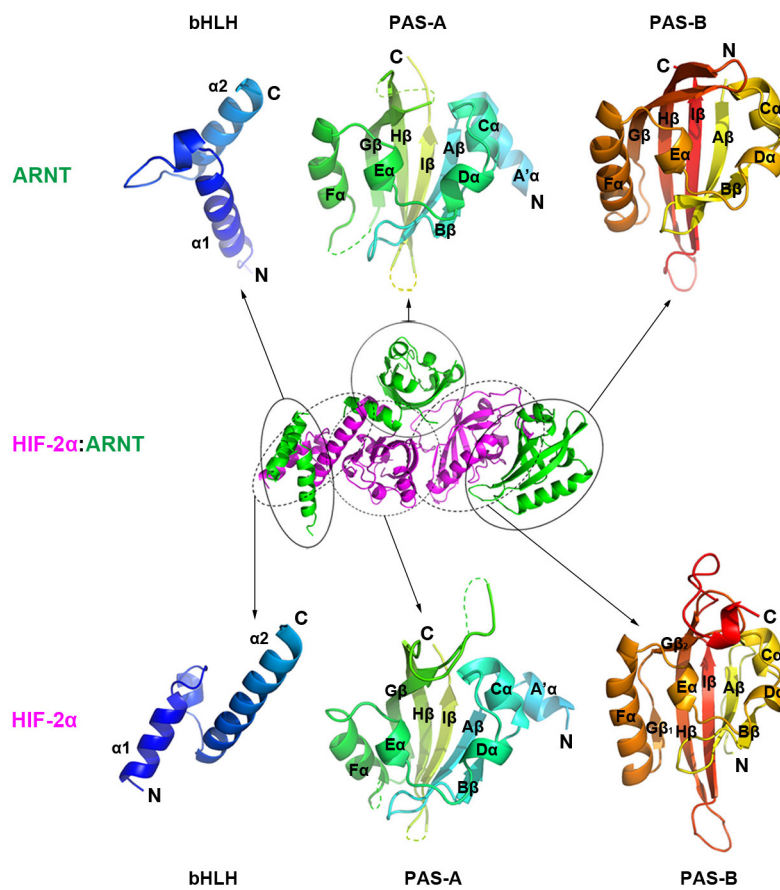
were seeded in 10 cm dishes and one day later transfected with 2 μ g pCMV-Tag4-HIF-2 α or HIF-1 α (WT or mutants) and 6 μ g pCMV-Tag1-ARNT (WT or mutants) plasmids using 16 μ l jetPRIME reagent (Polyplus-transfection). After overnight incubation, medium was refreshed with 50 μ M DP (2,2'-dipyridine, Sigma D216305). Another 24 h later cells were harvested and shortly sonicated in 700 μ l lysis buffer (1 \times TBS with 1 mM EDTA, 1% TRITON X-100 and 1 \times protease inhibitor cocktail) before quick spinning. After the protein concentration measurement for each sample, 40 μ g of supernatant was saved as input for western blot using monoclonal anti-Flag M2 antibody (Sigma F1804) and anti-Myc rabbit mAb (Cell Signaling 2278). Immunoprecipitation was performed with 1 mg of supernatant (diluted to 1 ml with lysis buffer) and 40 μ l of anti-Flag M2 affinity gel suspension (Sigma A2220) according to the manufacturer's instructions, followed by western blot using the anti-Myc antibody.

Compound binding assay. The intrinsic fluorescence signals of 10 nM acriflavine, proflavine (Sigma 131105) or tryaflavin (isolated from acriflavine by HPLC) were monitored using the BioTek Synergy 4 microplate reader (445 nm excitation/515 nm emission) in black 96-well plates. Purified HIF-2 α -ARNT or HIF-1 α -ARNT protein complexes were mixed with compounds in buffer containing 20 mM Tris pH 8.0 and 400 mM NaCl. The fluorescence quenching by gradually increased protein concentration was normalized to fraction bound, and the K_d values were calculated by fitting the curves in GraphPad Prism 6.

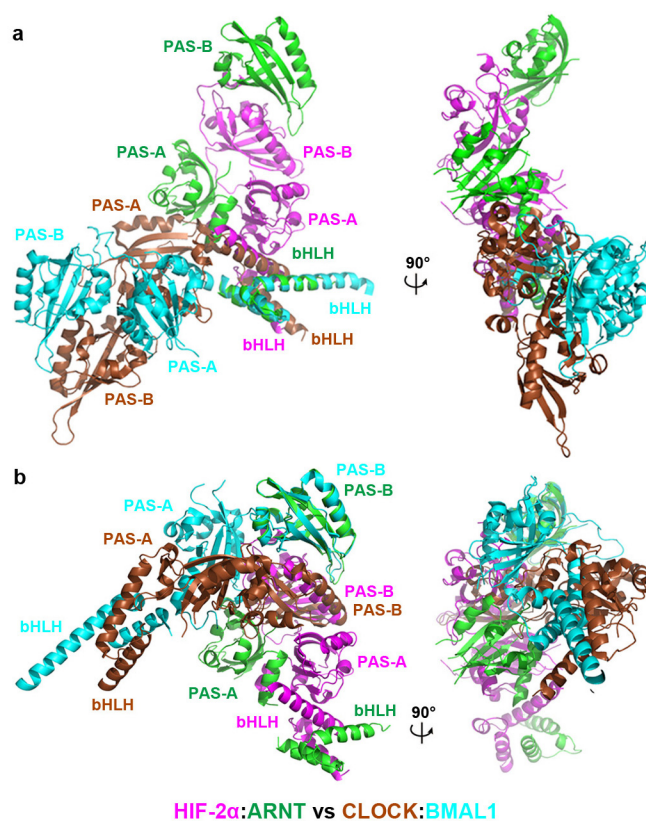
Fluorescence polarization HRE DNA binding assay. The fluoresceinated double-strand HRE DNA was prepared by annealing 6-FAM labelled forward strand (5'-GGCTGCGTACGTGCGGGTCGT-3') with the unlabelled reverse strand (5'-ACGACCCGCACGTACGCAGCC-3') in the buffer consisting of 20 mM Tris pH 7.5, 1 mM EDTA and 2 mM MgCl₂. For the binding assay, 2 nM DNA was incubated with purified complex proteins for 30 min, and final protein concentrations were varied by serial dilution in binding buffer (20 mM Tris pH 8.0, 50 mM NaCl). The fluorescence polarization signals were recorded using black 96-well plates on FlexStation 3 (Molecular Devices). The data were converted to fluorescence anisotropy, and K_d values were calculated as above.

- Wu, D., Potluri, N., Kim, Y. & Rastinejad, F. Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Mol. Cell. Biol.* **33**, 4346–4356 (2013).
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution from diffraction images to an initial model in minutes. *Acta Crystallogr. D* **62**, 859–866 (2006).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCr* **1**, 213–220 (2014).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–W118 (2006).
- Abaan, O. D. *et al.* The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology. *Cancer Res.* **73**, 4372–4382 (2013).
- Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nature Genet.* **44**, 694–698 (2012).
- Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genet.* **45**, 860–867 (2013).
- Seo, J. S. *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119 (2012).
- Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).





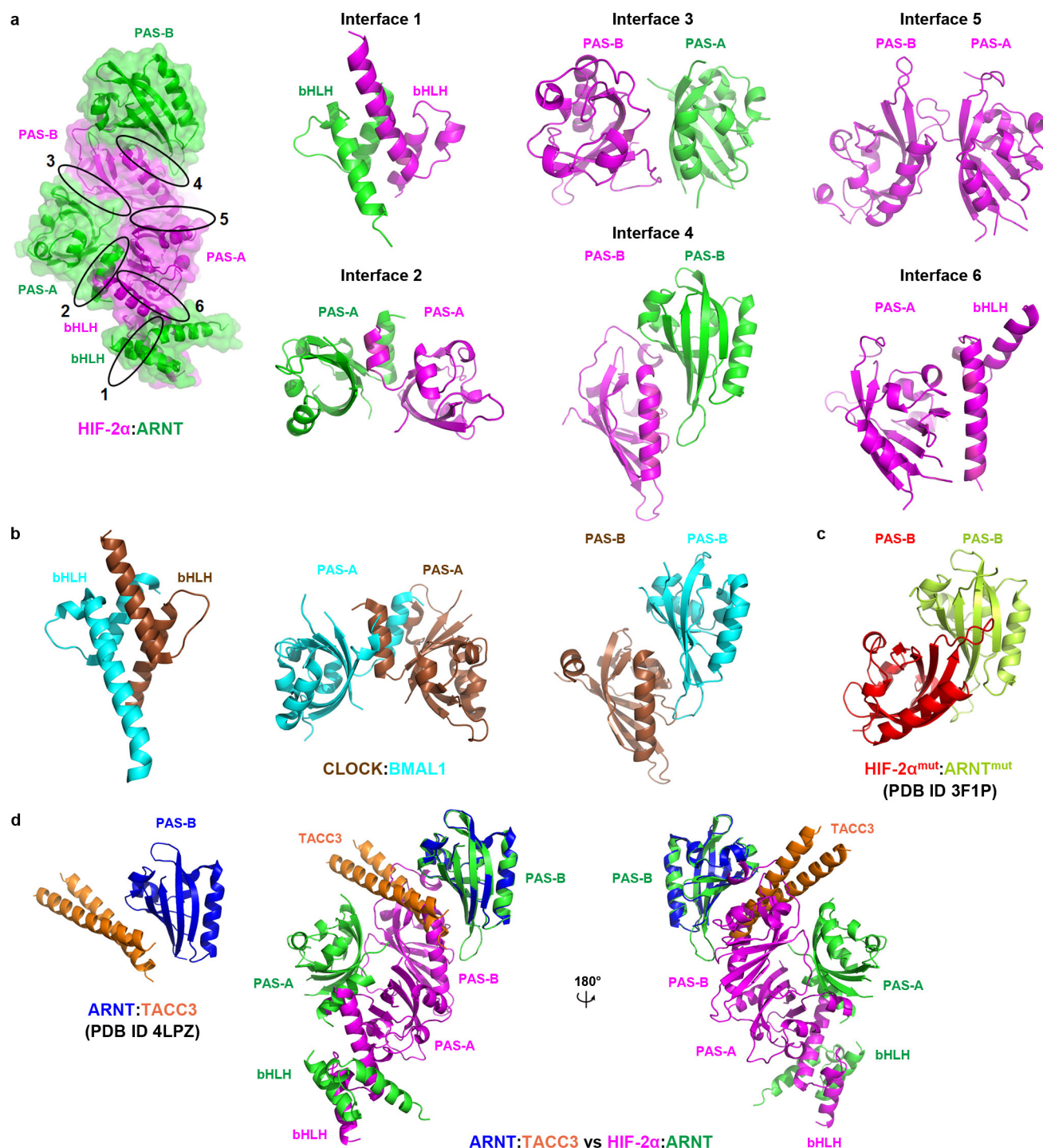
Extended Data Figure 2 | Structure of the HIF-2 α -ARNT complex in cartoon mode. Overall structure of HIF-2 α -ARNT complex (centre), and individual domains of ARNT (top) and HIF-2 α (bottom) with their secondary structures labelled.



HIF-2α:ARNT vs CLOCK:BMAL1

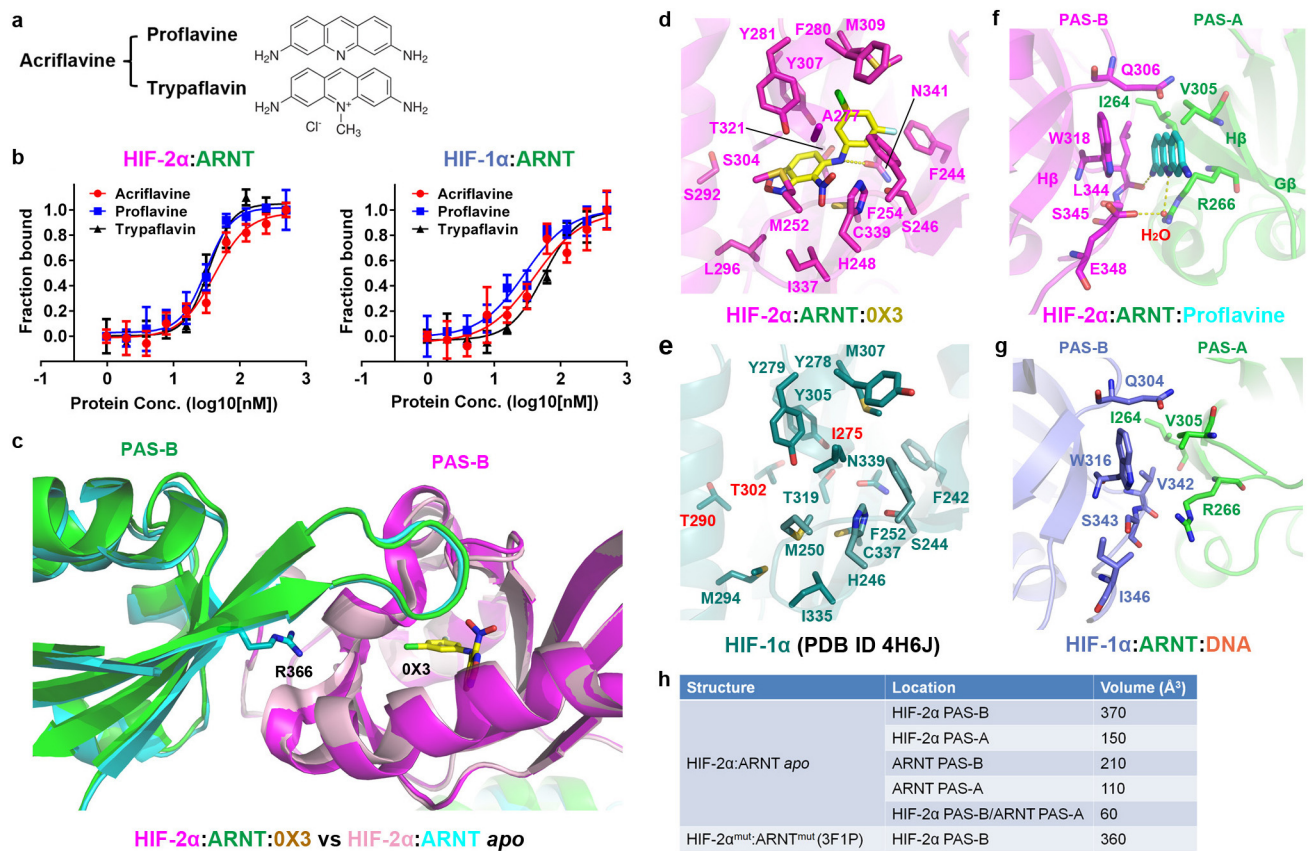
Extended Data Figure 3 | Comparison of the overall structures of HIF-2α–ARNT and CLOCK–BMAL1 complexes. Two complexes are superposed by aligning the bHLH domains (a) or PAS-B domains (b) of ARNT and

BMAL1, respectively. Colours used in labels match those used in figures for the same components.



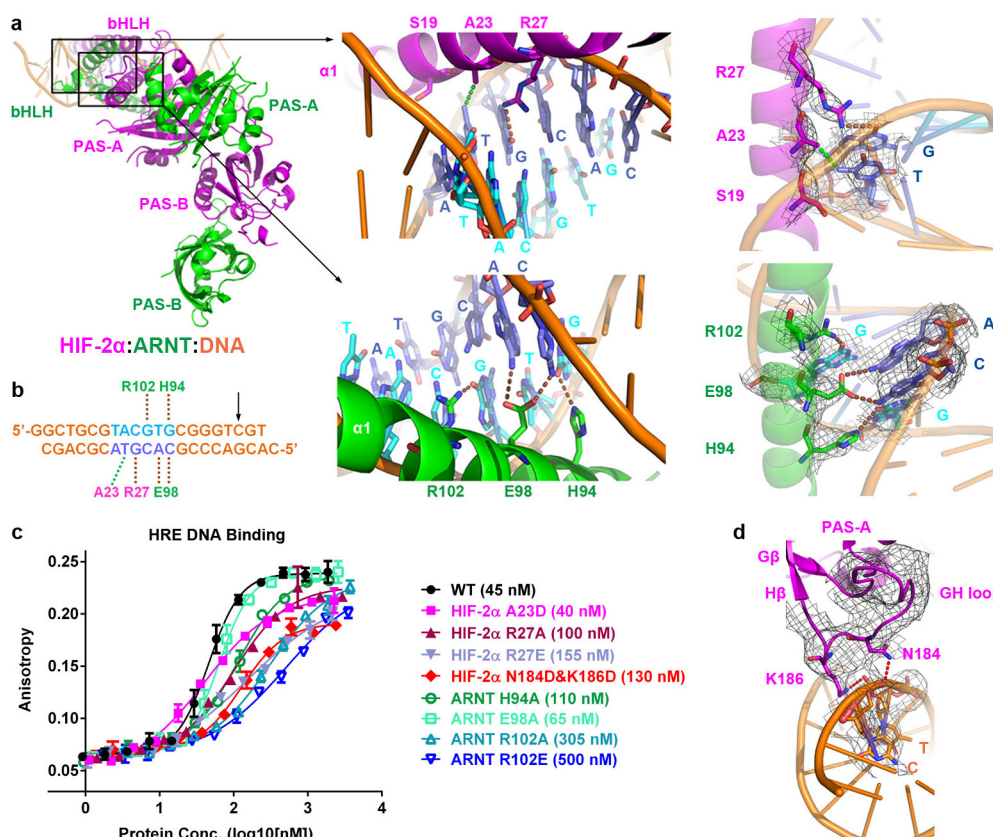
Extended Data Figure 4 | Domain interfaces of the HIF-2 α -ARNT complex and the comparison with other structures. **a**, Overall structure of HIF-2 α -ARNT complex with domain interfaces numbered (left), and the spatial arrangement of each interface (right). **b**, The dimer interfaces formed by equivalent domains in the CLOCK-BMAL1 complex, corresponding to the interfaces 1, 2 and 4 of HIF-2 α -ARNT heterodimer. **c**, The β -sheet-mediated antiparallel interface of isolated HIF-2 α -ARNT PAS-B complex with

mutations. **d**, The crystal structure showing the interaction between ARNT and TACC3 (left), and its superimpositions with HIF-2 α -ARNT complex through the PAS-B domain of ARNT in two views (right). Colours used in labels match those used in figures for the same components. The binding position of TACC3 peptide with respect to ARNT PAS-B cannot be fully accommodated in our quaternary structure, as some steric clashes would result involving TACC3 and the HIF-2 α PAS-B domain.



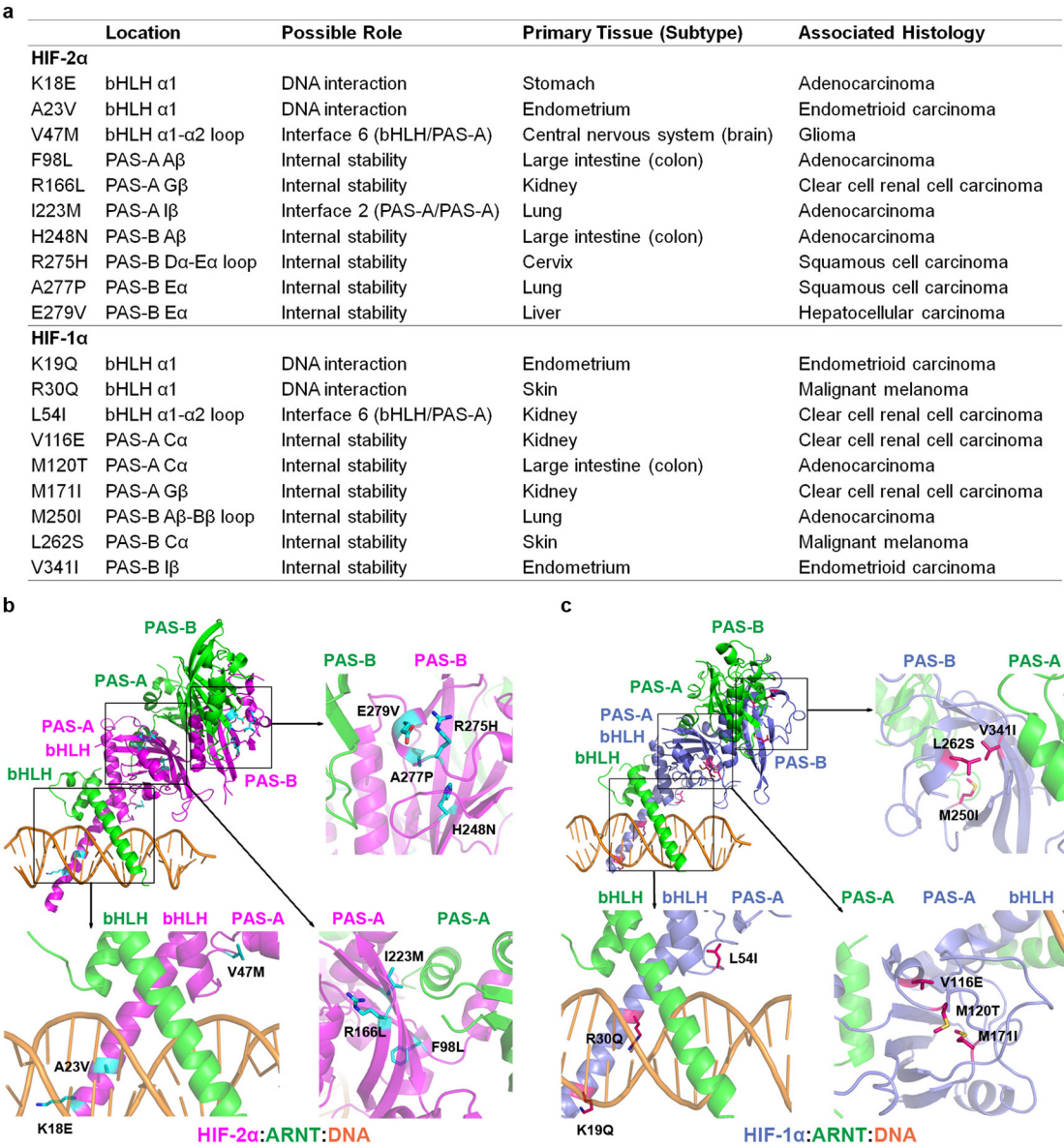
Extended Data Figure 5 | Ligand binding sites in the HIF-2 α -ARNT complex. **a**, Acriflavine is a mixture of proflavine and trypaflavin. **b**, Binding tests of acriflavine, proflavine and trypaflavin to protein complexes of HIF-2 α -ARNT (left) and HIF-1 α -ARNT (right). The calculated K_d values for acriflavine, proflavine and trypaflavin were 41 nM, 31 nM and 34 nM for HIF-2 α -ARNT, and 40 nM, 29 nM and 56 nM for HIF-1 α -ARNT, respectively. Representative data of at least two experiments, shown as mean \pm s.d. from three technical replicates. **c**, A proposed mechanism by which OX3 binding can destabilize the HIF-2 α -ARNT heterodimer. Shown is the proximity of OX3 within the PAS-B domain of HIF-2 α and R366 from the PAS-B domain of

ARNT. OX3 binding can potentially influence domain-domain interactions mediated by R366. We showed in Fig. 2b that R366 is a highly sensitive site for maintaining the stability of HIF-2 α -ARNT heterodimer. OX3 binding to the PAS-B domain of HIF-2 α could further influence interfaces 3, 4 and 5 (shown in Extended Data Fig. 4a). **d-g**, OX3 (**d**) binds at the inner pocket of HIF-2 α PAS-B domain, while proflavine (**f**) binds at the interfacial pocket formed by ARNT PAS-A and HIF-2 α PAS-B domains; corresponding positions in the structures of HIF-1 α PAS-B (**e**) and HIF-1 α -ARNT-DNA complex (**g**) are also shown. The yellow dotted lines represent hydrogen bonds. **h**, Pocket volumes calculated with CASTp program⁵⁸ at default 1.4 \AA probe radius.



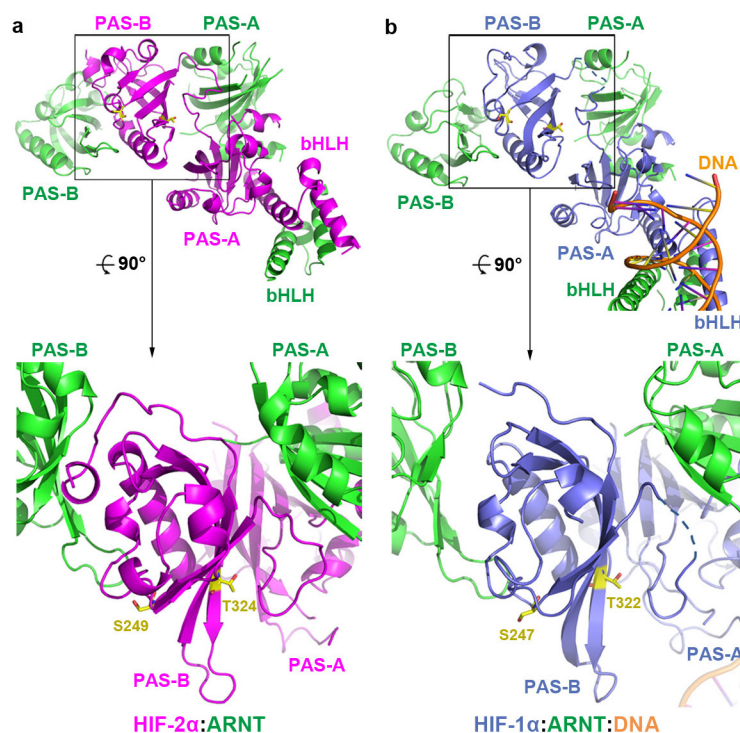
Extended Data Figure 6 | Interactions between the HIF-2 α -ARNT complex and DNA. **a**, Recognition of the HRE site by the bHLH domains. Overall structure of HIF-2 α -ARNT-DNA with each domain labelled (left); and detailed interactions between the core HRE site (blue) and the bHLH domains of HIF-2 α (upper right) or ARNT (lower right) are shown, with grey meshes showing $2F_o - F_c$ electron density contoured at 0.8σ . Hydrogen bonds (2.5–3.5 Å) are indicated by the brown dotted lines, while hydrophobic contact (3.6 Å) is shown by the green one. **b**, Schematic recognition diagram of HIF-2 α -ARNT to the HRE sequence (blue) on DNA. The brown and green dotted lines represent hydrogen bonds and hydrophobic contact, respectively. The black arrow indicates the nucleotide interacting with residues N184 and K186

from HIF-2 α PAS-A domain. Additional basic residues from HIF-2 α and ARNT bHLH domains (HIF-2 α K16, K18, R20, R24, R26 and ARNT R91, R99, R101) that can interact with DNA (mainly through the phosphate backbone) are also labelled in Extended Data Fig. 9. **c**, HRE DNA binding assay of HIF-2 α -ARNT protein complex in wild type (WT) or point-mutated forms using fluorescence polarization. Representative data of at least two experiments are shown as mean \pm s.d. from three technical replicates. Calculated approximate K_d values are shown in parentheses. **d**, The interactions between HIF-2 α PAS-A domain and DNA, with the GH loop (including residues N184 and K186) and interacting nucleotides meshed by $2F_o - F_c$ map at 0.8σ .



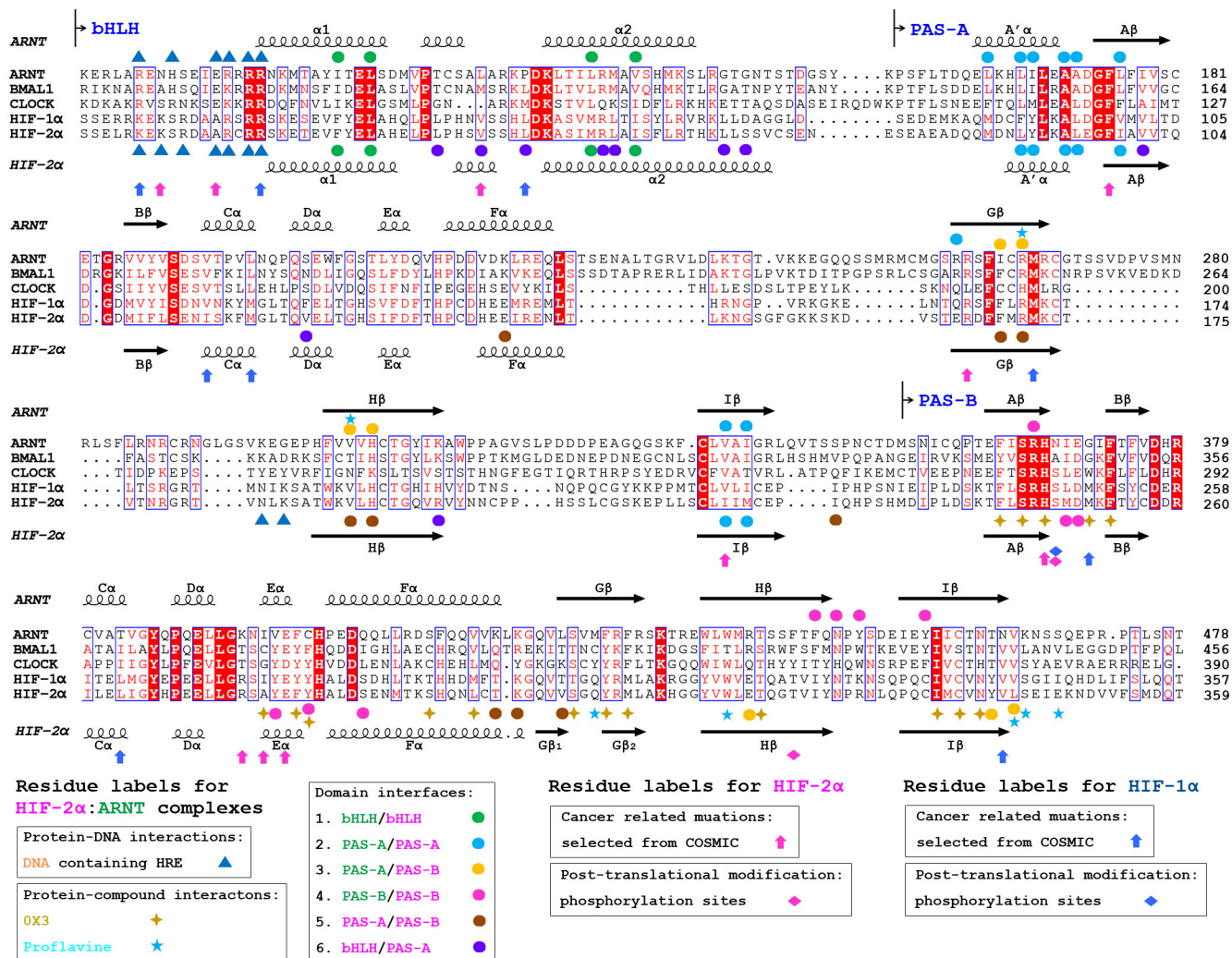
Extended Data Figure 7 | Locations of cancer-related missense mutations on the HIF-α-ARNT heterodimers. **a**, Detailed information of the selected cancer-related mutations in HIF-2α and HIF-1α. The information about tissue and histology was adopted from the COSMIC database⁴⁸ and other

publications^{45,59–62}. **b, c**, Spatial distribution of the HIF-2α (**b**) and HIF-1α (**c**) mutations in the heterodimers. Arrows point to close-up views of several regions in each heterodimer. The sequence positions of these mutations are also labelled in Extended Data Fig. 9.



Extended Data Figure 8 | Spatial positions of phosphorylation sites at the HIF- α PAS-B domains in the context of HIF- α -ARNT complexes. The phosphorylation sites T324 of HIF-2 α (a) and S247 of HIF-1 α (b) are shown as yellow sticks. Their corresponding residues T322 of HIF-1 α and S249 of HIF-2 α are also shown. Colours used in labels match those used in figures for the same components. HIF-2 α is phosphorylated at residue T324 by protein kinase D1, but the equivalent residue in HIF-1 α (T322) cannot be phosphorylated⁴⁹.

The differential positioning of their threonine residues next to a non-conserved loop between PAS-A and PAS-B domains may explain why they cannot be similarly phosphorylated. Casein kinase 1 (CK1) can phosphorylate HIF-1 α at S247 (ref. 50). We predict that the equivalent residue S249 in HIF-2 α may also be the target of CK1 (ref. 9), since the local environments for these residues are indistinguishable, with both residues being solvent accessible.



Extended Data Figure 9 | Comparison of mouse ARNT, BMAL1, CLOCK, HIF-1α and HIF-2α proteins. Sequence alignment of these five proteins includes the bHLH, PAS-A and PAS-B domains. Alignment was conducted by ClustalW2⁶³ and then processed by ESPript 3.0⁶⁴. The secondary structures of each domain of HIF-2α-ARNT complex, and residues involved in domain interfaces, protein-DNA or protein-compound interactions, are differently

labelled above (for ARNT) or below (for HIF-2α) the sequences. In addition, the HIF-2α (magenta) or HIF-1α (blue) residues with cancer-related mutations (mainly selected from the COSMIC database⁴⁸) and phosphorylation sites at the PAS-B domains are also labelled, further below the secondary structures of HIF-2α.

Extended Data Table 1 | Data collection and refinement statistics

	HIF-2 α :ARNT <i>apo</i>	HIF-2 α :ARNT: OX3	HIF-2 α :ARNT: Proflavine	HIF-2 α :ARNT: DNA	HIF-1 α :ARNT: DNA
Data collection					
Space group	P 1	P 1	P 1	P 43	P 43
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	49.0, 76.5, 98.8	49.1, 76.3, 98.0	48.9, 76.4, 98.3	65.7, 65.7, 245.9	66.7, 66.7, 243.6
α , β , γ (°)	90.1, 90.2, 73.1	89.9, 90.0, 73.2	89.8, 89.6, 73.2	90, 90, 90	90, 90, 90
Resolution (Å)	50.0–2.35 (2.39–2.35)*	50.0–2.87 (2.92–2.87)	50.0–2.80 (2.85–2.80)	50.0–3.60 (3.66–3.60)	50.0–3.90 (3.97–3.90)
<i>R</i> _{sym} or <i>R</i> _{merge}	6.1 (70.0)	12.1 (74.7)	5.4 (69.3)	9.7 (72.2)	4.5 (79.0)
<i>I</i> / σ <i>I</i>	11.8 (1.8)	8.9 (1.3)	17.0 (1.6)	12.9 (2.1)	24.7 (1.8)
Completeness (%)	98.4 (97.4)	97.9 (91.1)	98.5 (97.6)	99.5 (99.5)	99.1 (100)
Redundancy	2.6 (2.5)	2.5 (2.2)	2.4 (2.2)	3.7 (3.5)	3.8 (3.9)
Refinement					
Resolution (Å)	49.4–2.36 (2.40–2.36)	49.0–2.87 (2.96–2.87)	49.2–2.79 (2.88–2.79)	32.9–3.59 (3.95–3.59)	29.8–3.90 (4.11–3.90)
No. reflections	55769 (2688)	30376 (2510)	31508 (1754)	12131 (2992)	9578 (1356)
<i>R</i> _{work} / <i>R</i> _{free}	20.7/22.8 (33.9/37.2)	22.2/25.3 (34.9/39.7)	18.8/21.5 (24.8/29.3)	25.4/32.3 (30.3/35.2)	26.2/34.2 (38.7/43.9)
No. atoms					
Protein/DNA	8924	8695	8827	5624	5042
Ligand	0	21	16	0	0
Water	139	0	22	0	0
B-factors					
Protein/DNA	64.0	90.4	72.6	133.3	194.0
Ligand	-	103.8	80.0	-	-
Water	60.6	-	76.8	-	-
R.m.s deviations					
Bond lengths (Å)	0.011	0.012	0.010	0.004	0.005
Bond angles (°)	1.40	1.59	1.53	0.88	0.84

One crystal was used for each structure.

*Highest resolution shells are shown in parentheses.

The ubiquitin kinase PINK1 recruits autophagy receptors to induce mitophagy

Michael Lazarou^{1†*}, Danielle A. Sliter^{1*}, Lesley A. Kane^{1*}, Shireen A. Sarraf¹, Chunxin Wang¹, Jonathon L. Burman¹, Dionisia P. Sideris¹, Adam I. Fogel¹ & Richard J. Youle¹

Protein aggregates and damaged organelles are tagged with ubiquitin chains to trigger selective autophagy. To initiate mitophagy, the ubiquitin kinase PINK1 phosphorylates ubiquitin to activate the ubiquitin ligase parkin, which builds ubiquitin chains on mitochondrial outer membrane proteins, where they act to recruit autophagy receptors. Using genome editing to knockout five autophagy receptors in HeLa cells, here we show that two receptors previously linked to xenophagy, NDP52 and optineurin, are the primary receptors for PINK1- and parkin-mediated mitophagy. PINK1 recruits NDP52 and optineurin, but not p62, to mitochondria to activate mitophagy directly, independently of parkin. Once recruited to mitochondria, NDP52 and optineurin recruit the autophagy factors ULK1, DFCP1 and WIPI1 to focal spots proximal to mitochondria, revealing a function for these autophagy receptors upstream of LC3. This supports a new model in which PINK1-generated phospho-ubiquitin serves as the autophagy signal on mitochondria, and parkin then acts to amplify this signal. This work also suggests direct and broader roles for ubiquitin phosphorylation in other autophagy pathways.

Selective autophagy clears intracellular pathogens and mediates cellular quality control by engulfing cargo into autophagosomes and delivering it to lysosomes for degradation. Autophagy receptors bind ubiquitinated cargo and LC3-coated phagophores to mediate autophagy^{1,2}. Damaged mitochondria are removed by autophagy after activation of the kinase PINK1 and the E3 ubiquitin ligase parkin (encoded by *PARK2* in humans)^{3,4}. After the loss of mitochondrial membrane potential or the accumulation of misfolded proteins, PINK1 is stabilized on the outer mitochondrial membrane³, where it phosphorylates ubiquitin at Ser65 to activate parkin ubiquitin ligase activity^{5–7}. Although the autophagy receptors p62 and optineurin (OPTN) have been shown to bind ubiquitin chains on damaged mitochondria, their roles, and the roles of the other autophagy receptors in mediating mitophagy, are unclear^{8–11}.

Autophagy receptors in mitophagy

To clarify autophagy receptor function during mitophagy, genome editing was used to knock out five autophagy receptors (designated pentaKO) in HeLa cells, which do not express endogenous parkin. DNA sequencing (Supplementary Table 1) and immunoblotting of TAX1BP1, NDP52 (also known as CALCOCO2), NBR1, p62 (SQSTM1) and OPTN (Fig. 1a, lane 6) confirmed their knockout. We analysed mitophagy in pentaKOs by measuring the degradation of cytochrome C oxidase subunit II (COXII), a mitochondrial DNA (mtDNA)-encoded inner membrane protein, after mitochondrial damage with oligomycin and antimycin A. After oligomycin and antimycin A treatment, COXII was degraded in wild-type cells expressing parkin, but not in pentaKOs or ATG5 knockout HeLa cells, indicating a block in mitophagy (Fig. 1b, c, Supplementary Table 1 and Extended Data Fig. 1a). As a second indicator of mitophagy, mtDNA nucleoids were quantified by immunofluorescence (Extended Data Fig. 1b). After 24 h of oligomycin/antimycin A treatment, wild-type cells were nearly devoid of mtDNA, whereas mtDNA

was retained in pentaKOs and ATG5 knockouts (Fig. 1d, e). Parkin translocated to mitochondria (Extended Data Fig. 1c), and MFN1 and TOM20 (also known as TOMM20) were degraded via the proteasome comparably in wild-type and pentaKOs (Fig. 1b and Extended Data Fig. 1d). mtDNA nucleoids clump after oligomycin and antimycin A treatment in ATG5 knockout cells, but not in pentaKOs, consistent with a reported role of p62 (refs 10, 11).

The five endogenous receptors in wild-type cells (Extended Data Fig. 1c) and each receptor re-expressed in pentaKOs (Extended Data Fig. 1e, f) translocated to mitochondria after oligomycin/antimycin A treatment. However, in pentaKOs, only green fluorescent protein (GFP)-tagged NDP52, GFP-OPTN and, to a lesser extent, GFP-TAX1BP1 rescued mitophagy (Fig. 1f, g). Another recently reported autophagy receptor, TOLLIP¹², neither recruited to mitochondria nor rescued mitophagy after oligomycin/antimycin A treatment (Extended Data Fig. 1g–i).

We generated OPTN and NDP52 single knockout, NDP52/OPTN double knockout (DKO), and NDP52/OPTN/TAX1BP1 triple knockout cell lines (Supplementary Table 1 and Fig. 1a), and found no compensatory change in the expression of the remaining receptors. NDP52 or OPTN knockout alone caused no defect in mitophagy, whereas the NDP52/OPTN DKO and to a greater extent the NDP52/OPTN/TAX1BP1 triple knockout inhibited mitophagy (Fig. 2a–d and Extended Data Fig. 2a, b). The robust mitophagy observed in OPTN knockouts contrasts with a report indicating loss of mitophagy using RNA interference (RNAi)-mediated knockdown of OPTN in HeLa cells⁹. Although NDP52 and OPTN redundantly mediate mitophagy, they function non-redundantly in xenophagy¹³. Their expression levels in human tissues indicate that OPTN or NDP52 may function more prominently in different tissues (Extended Data Fig. 2c).

Mutations in autophagy receptors can lead to diseases such as primary open angle glaucoma (OPTN; E50K)¹⁴, amyotrophic lateral

¹Biochemistry Section, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA. [†]Present address: Department of Biochemistry and Molecular Biology, Monash University, Clayton, Melbourne 3800, Australia.

*These authors contributed equally to this work.

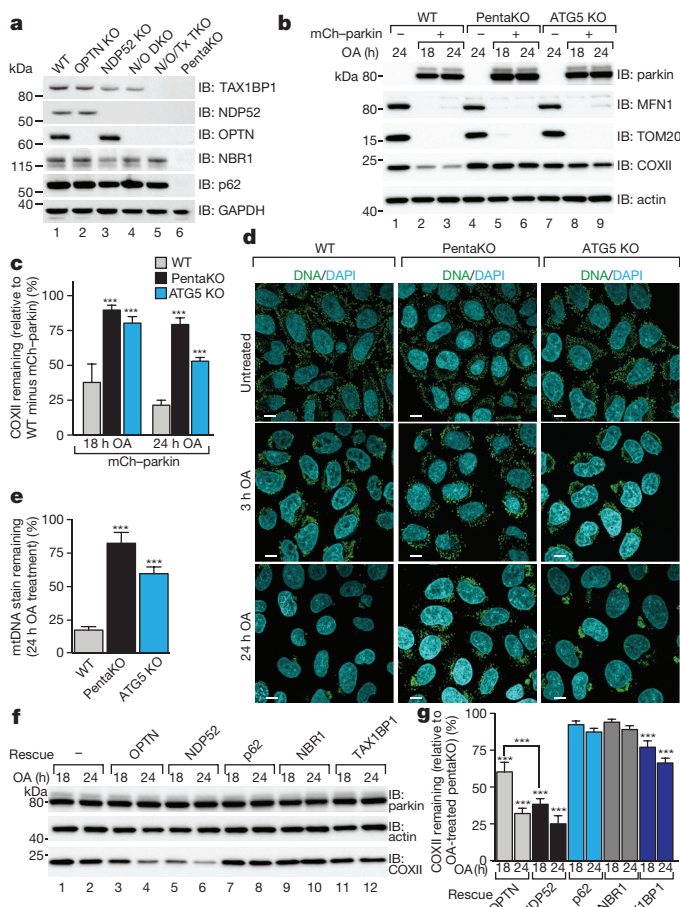


Figure 1 | Identifying autophagy receptors required for PINK1/parkin mitophagy. **a**, Wild-type (WT), OPTN knockout (KO), NDP52 knockout, NDP52/OPTN DKO (N/O DKO), NDP52/OPTN/TAX1BP1 triple knockout (N/O/Tx TKO), and pentaKO (NDP52/OPTN/TAX1BP1/NBR1/p62) HeLa cells were confirmed by immunoblotting (IB). kDa, kilodaltons. **b**, **c**, Cells as indicated with or without mCherry (mCh)-parkin were analysed by immunoblotting (**b**), and COXII levels were quantified (**c**). **d**, **e**, Representative images of mCh-parkin-expressing wild-type, pentaKO and ATG5 knockout cells immunostained to label mtDNA (green) (**d**), and quantified for mitophagy (24 h oligomycin/antimycin A (OA)) (**e**). More than 75 cells were counted per sample. DAPI, 4',6-diamidino-2-phenylindole. Scale bars, 10 μ m. **f**, **g**, Lysates from pentaKO cells expressing mCh-parkin and GFP-tagged autophagy receptors were immunoblotted (**f**) and COXII levels were quantified (**g**). Data in **c**, **e** and **g** are mean \pm s.d. from three independent experiments. *** P < 0.001 (one-way analysis of variance (ANOVA)).

sclerosis (OPTN; E478G and Q398X, in which X denotes a stop codon)¹⁵ and Crohn's disease (NDP52; V248A)¹⁶. Defects in xenophagy occur when OPTN is mutated to block its phosphorylation by TANK-binding kinase 1 (TBK1; S177A) or its ubiquitin binding (D474N)^{13,17}. In pentaKOs, the UBA domain-disrupting mutants OPTN (Q398X), OPTN(D474N) and OPTN(E478G) (Extended Data Fig. 2d) failed to translocate to mitochondria (Extended Data Fig. 2e, f) or rescue mitophagy (Fig. 2e and Extended Data Fig. 2g). OPTN(S177A) weakly rescued mitophagy and minimally translocated to mitochondria, whereas OPTN(E50K) robustly translocated and substantially rescued mitophagy (Fig. 2e and Extended Data Fig. 2e–g). NDP52(V248A) fully recruited to mitochondria and rescued mitophagy, but a mutant lacking the zinc-finger ubiquitin-binding domains (NDP52(Δ ZF))¹⁸ did not (Extended Data Fig. 2h–k and Fig. 2f). Thus, ubiquitin binding by OPTN and NDP52 is necessary for mitophagy, and some disease-causing mutations prevent mitophagy.

TBK1 and OPTN cooperate in mitophagy

TBK1 phosphorylation of OPTN at Ser177 increases its association with LC3 during xenophagy¹³, and the OPTN(E50K) mutation increases TBK1 and OPTN binding¹⁹. TBK1 auto-phosphorylation at Ser172 is indicative of TBK1 activation²⁰ and occurs in a parkin-dependent manner after 3 h of oligomycin/antimycin A treatment, but only in cells expressing OPTN (Extended Data Fig. 3b, lanes 4 and 10). Prolonged oligomycin and antimycin A treatment induces moderate TBK1 phosphorylation in the absence of parkin, but still requires PINK1 (Extended Data Fig. 3c). To investigate TBK1 function during mitophagy, we generated TBK1 knockout, and TBK1/NDP52 and TBK1/OPTN DKO HeLa cells (Extended Data Fig. 3d and Supplementary Table 1). Parkin translocated to mitochondria in all cell lines, however, only TBK1/NDP52 DKO cells displayed defective mitophagy (Extended Data Fig. 3e and Fig. 2g–j). Mitophagy in TBK1/NDP52 DKO cells was rescued by wild-type TBK1 or phosphomimetic OPTN (OPTN(S177D)), but not by kinase-dead TBK1 (TBK1(K38M); Extended Data Fig. 3g–i). Thus, in the absence of NDP52, TBK1 is crucial for effective mitophagy via OPTN.

Ubiquitin phosphorylation in mitophagy

Although many autophagy receptors recruit to the mitochondria after parkin activation, it is unclear why only some of these function in mitophagy. Parkin-mediated mitophagy is driven by the PINK1 phosphorylation of Ser65 of both ubiquitin^{5–7,21,22} and the parkin UBL domain²³. Because Ser65 phospho-ubiquitin is structurally unique, it may differentially interact with ubiquitin-binding proteins²². To determine whether OPTN is directly recruited to phospho-ubiquitin on mitochondria, we conditionally expressed PINK1 on undamaged mitochondria¹⁰ in HeLa cells lacking parkin (Fig. 3a and Extended Data Fig. 4a). When PINK1 Δ 110–YFP-2 \times FKBP (in which PINK1 lacks its N-terminal targeting sequence and is fused with yellow fluorescent protein (YFP) and two tandem FKBP domains at the C terminus) is cytosolic, mCherry-tagged OPTN (mCherry–OPTN), mCherry–NDP52 and mCherry–p62 are also cytosolic (Extended Data Fig. 4b, c). After rapalog treatment, PINK1 Δ 110–YFP-2 \times FKBP localizes to mitochondria expressing the FKBP12/rapamycin-binding (FRB)-Fis1 targeting construct where ubiquitin on mitochondrial surface proteins²⁴ (Extended Data Fig. 4a) can then be phosphorylated^{5,21,25,26}. This recruits OPTN and NDP52 to mitochondria (Fig. 3a, b), but not p62 (Fig. 3b and Extended Data Fig. 4c). OPTN and NDP52 recruitment requires PINK1 kinase activity (Fig. 3a, b) and receptor–ubiquitin binding, as the OPTN(D474N) and NDP52(Δ ZF) mutants fail to recruit after rapalog treatment (Fig. 3b and Extended Data Fig. 4d, e). Therefore, the ubiquitin kinase activity of PINK1 recruits OPTN and NDP52 to mitochondria via ubiquitin-binding domains in the absence of parkin.

To determine whether the observed autophagy receptor recruitment to mitochondria in the absence of parkin can induce mitophagy, we developed a sensitive fluorescence-activated cell sorting (FACS)-based mitophagy assay. We expressed mitochondrial-targeted mKeima (mt-mKeima, see Methods) in wild-type and pentaKO cells also expressing mitochondrial FRB-Fis1 and PINK1 Δ 110–YFP-2 \times FKBP. mt-mKeima engulfment into lysosomes results in a spectral shift owing to low pH. Only 1% (range 0.89–1.15) of wild-type or pentaKO cells display mitophagy when PINK1 is cytosolic. However, when PINK1 is recruited to mitochondria with rapalog, mitophagy increases ~sevenfold in wild-type cells and ~eightfold with overexpressed OPTN (Table 1 and Extended Data Fig. 5a). PentaKO cells showed no increase in mitophagy after targeting PINK1 to mitochondria (Table 1 and Extended Data Fig. 5b). When rescued with Flag/haemagglutinin (HA)-tagged OPTN or NDP52, pentaKOs displayed an increase in mitophagy of more than fivefold or fourfold, respectively (Table 1 and Extended Data Fig. 5b–e). Rescue with Flag/HA-p62 or ubiquitin-binding mutants (OPTN(Q398X), OPTN(D474N) and NDP52 (Δ ZF)) failed to increase mitophagy above baseline, but

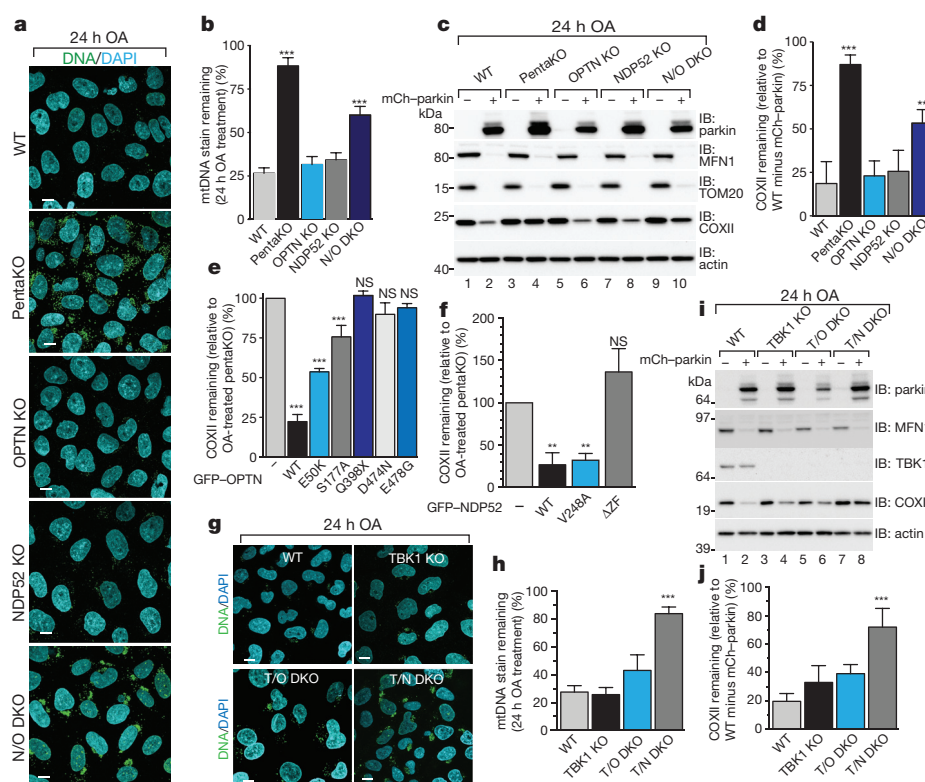


Figure 2 | OPTN and NDP52 are redundant in PINK1/parkin mitophagy.

a, b, Representative images of wild-type, pentaKO, OPTN knockout, NDP52 knockout and NDP52/OPTN DKO cells expressing mCh-parkin immunostained with anti-DNA (green) (**a**), and quantified for mitophagy (**b**). **c, d,** Cell lines from **a** were analysed by immunoblotting (**c**), and COXII levels were quantified (**d**). **e,** COXII levels quantified from pentaKO cells expressing mCh-parkin and rescued with wild-type or mutant GFP-OPTN (see Extended Data Fig. 2g for blots). **f,** COXII levels quantified from pentaKO cells expressing mCh-parkin rescued with wild-type or mutant GFP-NDP52

(see Extended Data Fig. 2k for blots). **g, h,** Representative images of wild-type, TBK1 knockout, TBK1/OPTN (T/O DKO) and TBK1/NDP52 (T/N DKO) HeLa cells expressing mCh-parkin and immunostained with anti-DNA (**g**), and quantification of mitophagy (**h**). **i, j,** Cells from **g** were immunoblotted (**i**) and COXII levels were quantified (**j**). Data in **b, d–f, h** and **i** are mean \pm s.d. from three independent experiments. ** $P < 0.005$, *** $P < 0.001$ (one-way ANOVA). NS, not significant. More than 75 cells were measured per confocal sample. Scale bars, 10 μ m. For untreated and mCh-parkin images of **a** and **g**, see Extended Data Fig. 3a and f, respectively.

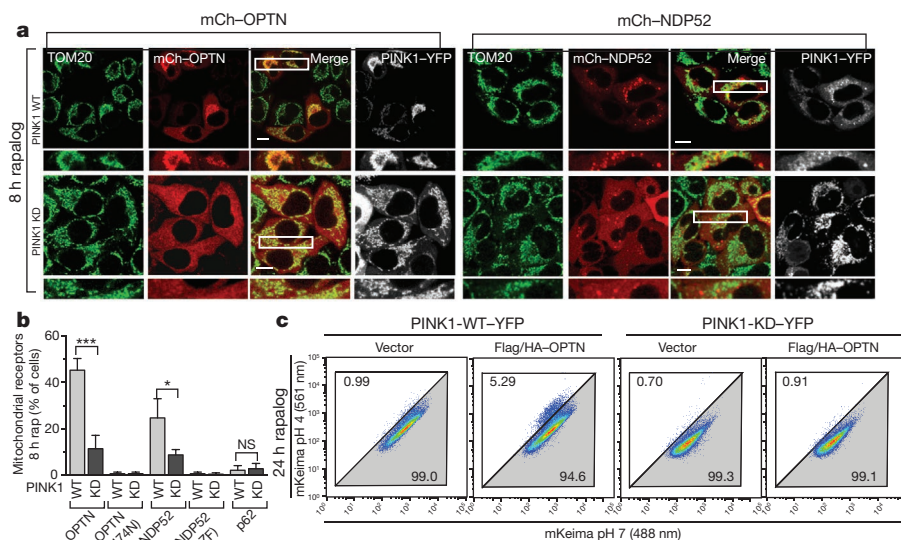


Figure 3 | PINK1 recruits OPTN and NDP52 independently of parkin to promote mitophagy. **a,** Representative images of HeLa cells expressing FRB-Fis1, wild-type or kinase-dead (KD) PINK1, PINK1 Δ 110-YFP-2 \times FKBP (PINK1-YFP), mCh-OPTN or mCh-NDP52, treated with rapalog. Scale bars, 10 μ m. **b,** Quantification of receptor translocation in cells from **a** and Extended Data Fig. 4c–e. More than 100 cells were counted per sample. **c,** Cells were

treated with rapalog and analysed by FACS for lysosomal-positive mt-mKeima. Representative data for wild-type or kinase-dead PINK1 Δ 110-YFP-2 \times FKBP in pentaKO cells with or without Flag/HA-OPTN expression. Data in **b** are mean \pm s.d. from three independent experiments. * $P < 0.05$, *** $P < 0.001$ (one-way ANOVA). For images of untreated cells from **a**, see Extended Data Fig. 4b.

Table 1 | Rapalog-induced mitophagy

Cell type	PINK1-YFP-2xFKBP	Receptor rescue	Receptor mutation	Parkin	Rap	Mean (%)	Fold change
WT	PINK1 WT	Vector	–	–	–	1.15	–
					+	8.37	7.28
		OPTN	WT	–	–	1.08	–
					+	8.59	7.95
PentaKO	PINK1 WT	Vector	–	–	–	0.93	–
					+	0.99	1.06
		OPTN	WT	–	–	0.89	–
					+	4.83	5.43
PentaKO	PINK1 WT	Vector	–	–	+	1.13	–
					+	5.27	4.66
	PINK1 KD	Vector	–	–	+	0.88	0.78
					+	0.76	0.67
PentaKO	PINK1 WT	Vector	–	–	+	0.93	–
		OPTN	WT	–	+	2.97	3.19
		OPTN	E50K	–	+	15.80	17.0
		OPTN	S177A	–	+	2.57	2.76
		OPTN	Q398X	–	+	0.87	0.78
		OPTN	D474N	–	+	0.76	0.67
	PINK1 WT	Vector	–	–	+	0.87	–
		NDP52	WT	–	+	3.50	3.98
		NDP52	V248A	–	+	2.80	3.18
		NDP52	ΔZF	–	+	0.47	0.53
	PINK1 WT	p62	WT	–	+	1.23	1.36
		OPTN	WT	–	–	1.10	–
		OPTN	WT	–	+	5.30	4.81
		OPTN	WT	+	–	1.43	–
		OPTN	WT	–	+	23.60	16.50

other mutants (OPTN(E50K), OPTN(S177A) and NDP52(V248A)) rescued mitophagy (Table 1 and Extended Data Fig. 5c–f). OPTN(E50K) and OPTN(S177A) restored mitophagy as well as or better than wild-type OPTN (Table 1), differing from their response in the presence of parkin (Fig. 2e), probably owing to the lack of robust TBK1 activation when parkin is absent (Extended Data Fig. 3b). Here, enhanced OPTN(E50K) binding to TBK1 (ref. 19) may become advantageous by allowing OPTN phosphorylation by TBK1 in the absence of parkin, thus improving mitophagy. In the absence of TBK1 activation, OPTN is probably not phosphorylated at Ser177 and thus is functionally similar to OPTN(S177A). Notably, the ubiquitin kinase activity of PINK1 is required, as kinase-dead PINK1 did not induce mitophagy (Table 1, Fig. 3c and Extended Data Fig. 5g). Parkin expression markedly increased mitophagy in Flag/HA–OPTN-expressing pentaKO cells (Table 1 and Extended Data Fig. 5h), supporting the model that PINK1-phosphorylated ubiquitin recruits receptors for mitophagy, and parkin ubiquitination of mitochondrial substrates amplifies this signal.

Comparing mitophagy induced by oligomycin/antimycin A treatment in wild-type relative to PINK1 knockout cells confirmed that endogenous PINK1 mediates mitophagy in the absence of parkin (Extended Data Fig. 6a, b). Re-expressing PINK1 in PINK1-knockout cells rescued oligomycin/antimycin-A-induced mitophagy (Extended Data Fig. 6c, d). Furthermore, mCherry–OPTN is recruited to mitochondria in the absence of parkin in a PINK1-dependent manner after prolonged exposure to oligomycin/antimycin A (Extended Data Fig. 6e, f).

Given that PINK1 ubiquitin kinase activity can recruit OPTN and NDP52, we investigated autophagy receptor binding to phosphomimetic (S65D) HA–ubiquitin in HeLa cells. Endogenous OPTN and NDP52 preferentially co-immunoprecipitate with HA–ubiquitin(S65D) (Extended Data Fig. 7a). Conversely, p62 was present at equal levels in all co-immunoprecipitates (Extended Data Fig. 7a). Ubiquitin-modified and unmodified forms of OPTN and NDP52 were present in co-immunoprecipitates, and HA–ubiquitin(S65D) induced or preserved this modification

(Extended Data Fig. 7a). Co-immunoprecipitated samples treated with the deubiquitinase USP2 removed the ubiquitin-modified bands on OPTN and NDP52, yet OPTN and NDP52 retained HA–ubiquitin(S65D) binding (Extended Data Fig. 7b). The binding of endogenous receptors in HeLa cell cytosol to *in vitro* phosphorylated strep-tactin-tagged ubiquitin (Extended Data Fig. 7c) showed that OPTN, but not p62, bound better to phospho-ubiquitin (Extended Data Fig. 7d, e). However, recombinant glutathione S-transferase (GST)–OPTN did not bind better to *in vitro* phosphorylated Lys63-linked ubiquitin chains²⁷, indicating that OPTN may need additional factors or modification *in vivo* to bind Ser65-phosphorylated ubiquitin preferentially.

OPTN and NDP52 recruit upstream machinery

Autophagy receptors are thought to function primarily by bridging LC3 and ubiquitinated cargo^{1,2}. In mCherry–parkin wild-type cells, GFP–LC3B accumulated in distinct puncta adjacent to mitochondria after oligomycin/antimycin A treatment (Extended Data Fig. 8a). Although oligomycin and antimycin A also induced GFP–LC3B puncta in pentaKOs, they were fewer and not near mitochondria (Extended Data Fig. 8a). Conversely, GFP–LC3B in ATG5 knockouts was near mitochondria, but not in puncta (Extended Data Fig. 8a). LC3B lipidation is retained in pentaKOs, but lost in ATG5 knockouts (Extended Data Fig. 8b). This indicates that ATG5 is activated downstream of PINK1, but independently of autophagy receptors, and that LC3 mitochondrial localization and lipidation are independent steps of mitophagy.

OPTN and NDP52 interact with LC3B and LC3C, respectively, for *Salmonella* clearance^{13,28}. Beyond that, little is known about the specificity of LC3 family members towards autophagy receptors²⁹ or their involvement in mitophagy. We examined the recruitment of all LC3 and GABARAP family members to mitochondria in wild-type, pentaKO and NDP52/OPTN DKO cells. The oligomycin/antimycin-A-induced mitochondrial localization of GFP–LC3s in wild-type cells was absent in pentaKO cells, while only GFP–LC3B recruitment was inhibited in NDP52/OPTN DKOs (Fig. 4a and Extended

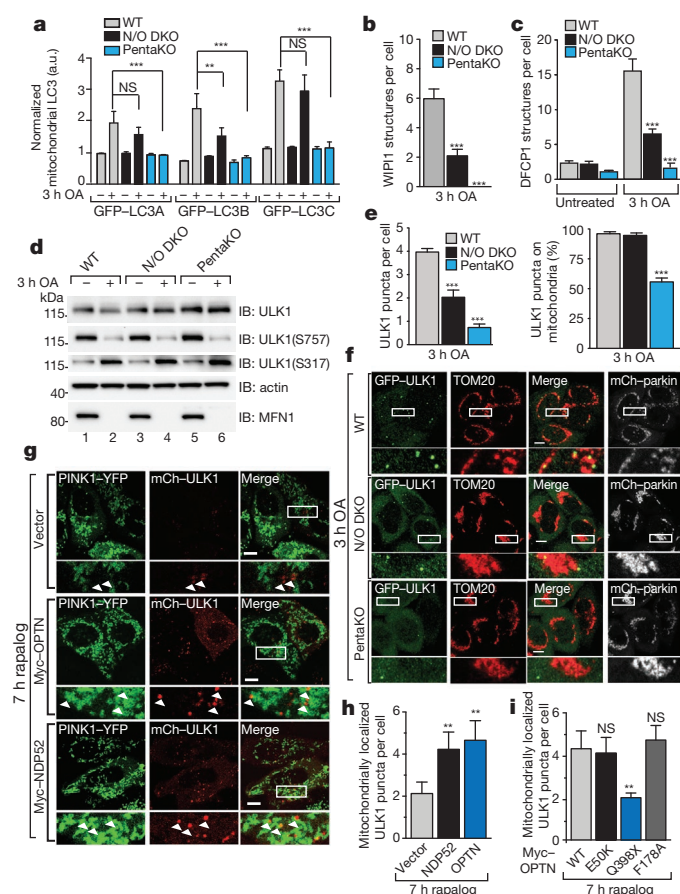


Figure 4 | Characterization of autophagy receptor function during mitophagy. **a–d**, mCh-parkin-expressing wild-type, NDP52/OPTN DKO and pentaKO cells were quantified for GFP-LC3A, GFP-LC3B and GFP-LC3C translocation to mitochondria (**a**), and for GFP-WIPI1 (**b**) or GFP-DFCP1 (**c**) structures per cell (>100 cells counted for each sample), or were immunoblotted using phospho-specific anti-S757 and anti-S317 ULK1 antibodies (**d**). **e**, mCh-parkin wild-type, NDP52/OPTN DKO and pentaKO cells stably expressing GFP-ULK1 were quantified for GFP-ULK1 puncta per cell (left), and the percentage of those puncta on mitochondria was determined (right). **f**, Representative cells from **e** were immunostained for TOM20 and GFP. **g**, PentaKO cells expressing FRB-Fis1, PINK1Δ110-YFP-2×FKBP, mCh-ULK1 and Myc-tagged receptors were treated with rapalog and then imaged live. **h**, Quantification of mitochondrial ULK1 puncta in **g**. **i**, Quantification of mitochondrial ULK1 puncta in pentaKO cells expressing FRB-Fis1, PINK1Δ110-YFP-2×FKBP, mCh-ULK1 and Myc-OPTN mutants, treated with rapalog then imaged live. Data in **a–c**, **e**, **h** and **i** are mean ± s.d. from three independent experiments. ** $P < 0.005$, *** $P < 0.001$ (one-way ANOVA). For live-cell quantification, >75 cells were counted in a blinded manner. Quantifications in **h** and **i** were performed after removal of outliers; see Methods for details. Scale bars, 10 μm. a.u., arbitrary units. See Extended Data Figs 8c, 9b, c and 10d for representative images of **a**, **b**, **c** and **i**, respectively. See Extended Data Fig. 9e for untreated samples of **f**, and Extended Data Fig. 10c for untreated images of **h** and **i**.

Data Fig. 8c). GFP-LC3C recruitment was inhibited in NDP52/OPTN/TAX1BP1 triple knockouts (Extended Data Fig. 8d, e), indicating that TAX1BP1 can recruit LC3C during mitophagy. GABARAP proteins did not recruit to mitochondria, indicating that they probably do not have a substantial role in mitophagy (Extended Data Fig. 9a).

We also examined the involvement of WIPI1 and DFCP1, two proteins that mediate phagophore biogenesis upstream of LC3 (ref. 30), in mitophagy. In wild-type cells, oligomycin/antimycinA-induced foci of both GFP-WIPI1 and GFP-DFCP1, mostly localized on or near mitochondria (Fig. 4b, c and Extended Data Fig. 9b, c). In NDP52/OPTN DKO cells, GFP-WIPI1 and GFP-DFCP1 foci were reduced, and

almost undetectable in pentaKO cells (Fig. 4b, c and Extended Data Fig. 9b, c). Despite this, the phosphorylation of beclin1 (ref. 31) was normal in both pentaKO and NDP52/OPTN DKO cells (Extended Data Fig. 9d), indicating that failure to recruit WIPI1 and DFCP1 was not due to a defective VPS34 (also known as PIK3C3) complex. GFP-DFCP1 recruitment in pentaKO cells was rescued by the expression of Flag/HA-OPTN or Flag/HA-NDP52, but not by Flag/HA-p62 (Extended Data Fig. 10a).

Although autophagy receptors are thought to function late in autophagy with LC3 (ref. 32), the deficit in WIPI1 and DFCP1 recruitment to mitochondria indicates a defect upstream in autophagosome biogenesis. ULK1 phosphorylation by AMPK at Ser317 and dephosphorylation at Ser757 (ref. 33), required for activation, occurs comparably in wild-type, NDP52/OPTN DKO and pentaKO cells (Fig. 4d). Despite this, ULK1 puncta appearance³⁴ after oligomycin and antimycin A treatment is diminished by half in the NDP52/OPTN DKO cells, and by more than 80% in pentaKO cells (Fig. 4e, f). Flag/HA-OPTN or Flag/HA-NDP52, but not Flag/HA-p62, rescued GFP-ULK1 localization in pentaKO cells (Extended Data Fig. 10b). Overall, these data indicate that NDP52 and OPTN recruit ULK1 to initiate mitophagy.

We next assessed whether ubiquitin phosphorylation, independent of parkin, is also sufficient to recruit ULK1 to mitochondria. Rescue of pentaKO cells expressing FRB-Fis1 and PINK1Δ110-YFP-2×FKBP with Myc-OPTN or Myc-NDP52 resulted in mitochondrial ULK1 puncta after rapalog treatment (Fig. 4g, h). Myc-OPTN(E50K) also rescued ULK1 recruitment to mitochondria, but mutant Myc-OPTN(Q398X), which is associated with amyotrophic lateral sclerosis, did not (Fig. 4i and Extended Data Fig. 10d). ULK1 recruitment was restored by Myc-OPTN(F178A) (Fig. 4i and Extended Data Fig. 10d), a mutation that disrupts OPTN association with LC3 (ref. 13), indicating that ULK1 recruitment is not through LC3 interaction and occurs upstream of LC3 (ref. 34). Taken together, our data show that PINK1 ubiquitin kinase activity is sufficient to recruit the autophagy receptors and upstream autophagy machinery to mitochondria to induce mitophagy.

Conclusions

Through the genetic knockout of five autophagy receptors, we have defined their relative roles in mitophagy and identified their unanticipated upstream involvement in autophagy recruitment machinery. p62 and NBR1 are dispensable for parkin-mediated mitophagy; OPTN and NDP52 are the primary, yet redundant, receptors. We also uncovered a new and more fundamental role for PINK1 in mitophagy: to induce mitophagy directly through phospho-ubiquitin-mediated recruitment of autophagy receptors. We propose that PINK1 generates the novel and essential signature (phospho-ubiquitin) on mitochondria to induce OPTN and NDP52 recruitment and mitophagy; parkin acts to increase this signal by generating more ubiquitin chains on mitochondria, which are subsequently phosphorylated by PINK1. Our findings clarify the role of parkin as an amplifier of the PINK1-generated mitophagy signal, phospho-ubiquitin, which can engage the autophagy receptors to recruit ULK1, DFCP1, WIPI1 and LC3 (see model in Extended Data Fig. 10e).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 January; accepted 10 July 2015.

Published online 12 August 2015.

1. Svenning, S. & Johansen, T. Selective autophagy. *Essays Biochem.* **55**, 79–92 (2013).
2. Stolz, A., Ernst, A. & Dikic, I. Cargo recognition and trafficking in selective autophagy. *Nature Cell Biol.* **16**, 495–501 (2014).
3. Narendra, D. P. *et al.* PINK1 is selectively stabilized on impaired mitochondria to activate parkin. *PLoS Biol.* **8**, e1000298 (2010).

4. Narendra, D., Tanaka, A., Suen, D. F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803 (2008).
5. Kane, L. A. *et al.* PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *J. Cell Biol.* **205**, 143–153 (2014).
6. Kazlauskaitė, A. *et al.* Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65. *Biochem. J.* **460**, 127–139 (2014).
7. Koyano, F. *et al.* Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* **510**, 162–166 (2014).
8. Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
9. Wong, Y. C. & Holzbaur, E. L. F. Optineurin is an autophagy receptor for damaged mitochondria in parkin-mediated mitophagy that is disrupted by an ALS-linked mutation. *Proc. Natl Acad. Sci. USA* **111**, E4439–E4448 (2014).
10. Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. p62/SQSTM1 is required for Parkin-induced mitochondrial clustering but not mitophagy; VDAC1 is dispensable for both. *Autophagy* **6**, 1090–1106 (2010).
11. Okatsu, K. *et al.* p62/SQSTM1 cooperates with Parkin for perinuclear clustering of depolarized mitochondria. *Genes Cells* **15**, 887–900 (2010).
12. Lu, K., Psakhye, I. & Jentsch, S. Autophagic clearance of polyQ proteins mediated by ubiquitin-Atg8 adaptors of the conserved CUET protein family. *Cell* **158**, 549–563 (2014).
13. Wild, P. *et al.* Phosphorylation of the autophagy receptor optineurin restricts Salmonella growth. *Science* **333**, 228–233 (2011).
14. Rezaie, T. *et al.* Adult-onset primary open-angle glaucoma caused by mutations in optineurin. *Science* **295**, 1077–1079 (2002).
15. Maruyama, H. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465**, 223–226 (2010).
16. Ellinghaus, D. *et al.* Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. *Gastroenterology* **145**, 339–347 (2013).
17. Mankouri, J. *et al.* Optineurin negatively regulates the induction of IFN β in response to RNA virus infection. *PLoS Pathog.* **6**, e1000778 (2010).
18. Thurston, T. L., Ryzhakov, G., Bloor, S., von Muhlen, N. & Randow, F. The TBK1 adaptor and autophagy receptor NDP52 restricts the proliferation of ubiquitin-coated bacteria. *Nature Immunol.* **10**, 1215–1221 (2009).
19. Morton, S., Hesson, L., Pegg, M. & Cohen, P. Enhanced binding of TBK1 by an optineurin mutant that causes a familial form of primary open angle glaucoma. *FEBS Lett.* **582**, 997–1002 (2008).
20. Larabi, A. *et al.* Crystal structure and mechanism of activation of TANK-binding kinase 1. *Cell Rep.* **3**, 734–746 (2013).
21. Ordureau, A. *et al.* Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Mol. Cell* **56**, 360–375 (2014).
22. Wauer, T. *et al.* Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J.* **34**, 307–325 (2015).
23. Kondapalli, C. *et al.* PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biol.* **2**, 120080 (2012).
24. Sarraf, S. A. *et al.* Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. *Nature* **496**, 372–376 (2013).
25. Shiba-Fukushima, K. *et al.* Phosphorylation of mitochondrial polyubiquitin by PINK1 promotes Parkin mitochondrial tethering. *PLoS Genet.* **10**, e1004861 (2014).
26. Okatsu, K. *et al.* Phosphorylated ubiquitin chain is the genuine Parkin receptor. *J. Cell Biol.* **209**, 111–128 (2015).
27. Ordureau, A. *et al.* Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. *Proc. Natl Acad. Sci. USA* **112**, 6637–6642 (2015).
28. von Muhlen, N. *et al.* LC3C, bound selectively by a noncanonical LIR motif in NDP52, is required for antibacterial autophagy. *Mol. Cell* **48**, 329–342 (2012).
29. Wild, P., McEwan, D. G. & Dikic, I. The LC3 interactome at a glance. *J. Cell Sci.* **127**, 3–9 (2014).
30. Lamb, C. A., Yoshimori, T. & Tooze, S. A. The autophagosome: origins unknown, biogenesis complex. *Nature Rev. Mol. Cell Biol.* **14**, 759–774 (2013).
31. Fogel, A. I. *et al.* Role of membrane association and Atg14-dependent phosphorylation in beclin-1-mediated autophagy. *Mol. Cell Biol.* **33**, 3675–3688 (2013).
32. Koyama-Honda, I., Itakura, E., Fujiwara, T. K. & Mizushima, N. Temporal analysis of recruitment of mammalian ATG proteins to the autophagosome formation site. *Autophagy* **9**, 1491–1499 (2013).
33. Kim, J., Kundu, M., Viollet, B. & Guan, K. L. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biol.* **13**, 132–141 (2011).
34. Itakura, E., Kishi-Itakura, C., Koyama-Honda, I. & Mizushima, N. Structures containing Atg9A and the ULK1 complex independently target depolarized mitochondria at initial stages of Parkin-mediated mitophagy. *J. Cell Sci.* **125**, 1488–1499 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Nezich and S. Banerjee in the Youle laboratory, C. Smith and the NINDS and NHLBI Flow Cytometry Core Facilities. This work was supported by the Intramural Research Program of the NIH, NINDS and the National Health and Medical Research Council (GNT1063781).

Author Contributions M.L., D.A.S., L.A.K. and R.J.Y. conceived the projects; M.L., D.A.S., L.A.K., S.A.S., C.W., D.P.S., A.I.F. and R.J.Y. designed experiments; M.L., D.A.S., L.A.K., S.A.S., C.W., J.L.B., D.P.S. and A.I.F. performed experiments; M.L., D.A.S., L.A.K. and R.J.Y. wrote the manuscript, and all authors contributed to editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.Y. (youler@ninds.nih.gov).

METHODS

Cell culture, antibodies and reagents. HEK293T, HeLa and PINK1 knockout cells³⁵ were cultured in DMEM (Life Technologies) supplemented with 10% (v/v) FBS (Gemini Bio Products), 10 mM HEPES (Life Technologies), 1 mM sodium pyruvate (Life Technologies), non-essential amino acids (Life Technologies) and GlutaMAX (Life Technologies). HeLa cells were acquired from the ATCC and authenticated by the Johns Hopkins GRCF Fragment Analysis Facility using STR profiling. All cells were tested for mycoplasma contamination bimonthly using the Plasmotest kit (InvivoGen). Transfection reagents used were: effectene (Qiagen), lipofectamine LTX (Life Technologies), avalanche-OMNI (EZ Biosystems), X-tremeGENE HP (Roche) and X-tremeGENE 9 (Roche).

The following rabbit monoclonal and polyclonal antibodies were used: beclin, pULK1(S317), pULK1(S757), TBK1, pTBK1(S172), NDP52, TAX1BP1, ATG5, actin, and haemagglutinin (Cell Signaling Technologies); GAPDH and LC3B (Sigma); ULK1 and TOM20 (Santa Cruz Biotechnology); OPTN (Proteintech); GFP (Life Technologies); pSer65 ubiquitin (Millipore) and MFN1 was generated previously³⁶. Mouse monoclonal antibodies used were: NBR1 and p62 (Abnova), COXII (Abcam), parkin (Santa Cruz Biotechnology), DNA (Progen Biotechnik) and ubiquitin (Cell Signaling). Chicken anti-GFP (Life Technologies) was also used. For catalogue numbers, see Supplementary Table 1. Human tissue panel blots were purchased (NOVUS Biologicals).

Generation of knockout lines using TALEN and CRISPR/Cas9 gene editing. To generate knockout cell lines, transcription activator-like effector nuclease (TALEN) and CRISPR guide RNAs (gRNAs) were chosen that targeted an exon common to all splicing variants of the gene of interest (listed in Supplementary Table 1). TALEN was used to generate the OPTN knockout HeLa cell line. The TALEN constructs were generated by sequential ligation of coding repeats into pcDNA3.1/Zeo-Talen(+63), as previously described^{37–39}. The CRISPR/Cas9 system generated previously⁴⁰ was used to knockout ATG5, NDP52, TAX1BP1, NBR1, p62 and TBK1. Oligonucleotides (Operon) containing CRISPR target sequences were annealed and ligated into AlflI-linearized gRNA vector (Addgene 41824)⁴⁰. For CRISPR/Cas9 gene editing, HeLa cells were transfected with gRNA constructs, hCAS9 (Addgene 41815) and pEGFP-C1 (Clontech). For TALEN gene editing, HeLa cells were transfected with OPTN TALEN constructs and pEGFP-C1. Two days after transfection, GFP-positive cells were sorted by FACS and plated in 96-well plates. Single colonies were expanded into 24-well plates before screening for depletion of the targeted gene product by immunoblotting. As a secondary screen of some knockout lines, genomic DNA was isolated from cells and the genomic regions of interest were amplified using PCR followed restriction enzyme digestion analysis (primers listed in Supplementary Table 1). Sequencing of targeted genomic regions of knockout lines was also conducted to confirm the presence of frameshifting indels in the genes of interest (Supplementary Table 1). To generate multiple gene knockout cell lines, parental cell lines were transfected sequentially with one or multiple gRNA constructs to generate desired knockout lines. Parental cell lines are outlined in Supplementary Table 1.

Cloning and generation of stable cell lines. pMXs-puro-GFP-WIP1 and pMXs-puro-GFP-DFC1 were a gift from N. Mizushima, and pMXs-IP-GFP-ULK1 was purchased from Addgene (38193). To generate pBMN-mEGFP-C1, mEGFP-C1 (Addgene 36412) was PCR amplified (together with the multiple cloning site) and cloned into pBMN-Z at BamHI/SalI sites using the Gibson Cloning kit (New England Biolabs) according to the manufacturer's instructions. The BamHI and SalI sites used to insert mEGFP-C1 were not regenerated. The following GFP-tagged plasmids were generated by PCR amplification of open reading frames followed by ligation into pBMN-mEGFP-C1: OPTN, NDP52, p62, TAX1BP1, NBR1, LC3A, LC3B, LC3C, GABARAP, GABARAPL1 and GABARAPL2. The Gateway Cloning (Invitrogen) system was used to generate GFP-, mCherry-, Myc- and Flag/HA-tagged constructs. In brief, TBK1, TBK1(K38M), NDP52, OPTN, p62, DFC1, WIP1 and ULK1 were cloned into pDONR2333. Mutations in complementary DNA sequences were introduced using PCR site-directed mutagenesis in the pDONR2333 vector, (sequences of mutagenesis primers used are available on request) then recombined into pHAGE-N-Flag/HA, pHAGE-N-GFP, pHAGE-N-mCherry and/or pDEST-N-myc using LR Clonase (Invitrogen) as per the manufacturer's protocol. All constructs generated in this study were verified by sequencing.

To generate stably transfected cell lines, retroviruses (for pBMN-mEGFP-C1 constructs, pBMN-mCherry-parkin, pBMN-puro-P2A-FRB-Fis1, pCHAC-mt-mKeima-IRES-MCS2) and lentiviruses (for pHAGE and pDEST constructs) were packaged in HEK293T cells. HeLa cells were transduced with virus for 24 h with $8 \mu\text{g ml}^{-1}$ polybrene (Sigma) then optimized for protein expression via selection (puromycin or blasticidin) or fluorescence sorting.

Translocation and mitophagy treatments. Cells were either left untreated or treated with 10 μM oligomycin (Calbiochem), 4 μM antimycin A (Sigma) in fresh

growth medium for different periods of time as indicated. Some experiments were performed with 10 μM carbonyl cyanide *m*-chlorophenyl hydrazine (CCCP) as indicated (Sigma-Aldrich). We chose to use oligomycin/antimycin A to depolarize mitochondria in most of our experiments, as they are specific mitochondrial respiratory complex inhibitors and less toxic. Long treatment time points of both oligomycin/antimycin A and CCCP were also supplemented with the apoptosis inhibitor 20 μM QVD (ApexBio) to prevent cell death.

Immunoblotting and phos-tag gels. HeLa cells seeded into six-well plates were either untreated or treated with 10 μM oligomycin (Calbiochem), 4 μM antimycin A (Sigma) and 20 μM QVD (ApexBio) in fresh growth medium for different periods of time as indicated. Cells were lysed in 1 \times LDS sample buffer (Life Technologies) supplemented with 100 mM dithiothreitol (DTT; Sigma) and heated to 99 °C with shaking for 7–10 min. Approximately 25–50 μg of protein per sample was separated on 4–12% Bis-Tris gels (Life Technologies) according to manufacturer's instructions, and then transferred to polyvinylidene difluoride membranes and immunoblotted using antibodies as indicated. To assess mitophagy, COXII quantification was conducted using ImageLab software (BioRad). For uncropped images of all immunoblots, see Supplementary Information.

To dephosphorylate samples, cells were collected as above and lysed in 1 \times NEB buffer 3 (New England Biolabs) supplemented with 1% Triton X-100 and passed through a 26.5-gauge needle. Calf intestinal phosphatase (CIP; New England Biolabs) was added to half of the cell lysate and the other half was used as an untreated control. Both samples were incubated for 1 h at 37 °C and analysed by SDS-PAGE and immunoblotting.

To analyse beclin phosphorylation, lysates were prepared in sample buffer lacking EDTA and run on 8% Tris-glycine gels containing 20 μM phos-tag (Wako) and 40 μM MnCl₂ as described previously³¹. Gels lacking phos-tag were run simultaneously as a negative control. Electrophoresis and western transfer were carried out using standard protocols with the exception that phos-tag gels were incubated in 10 mM EDTA for 10 min to remove excess Mn²⁺ before transfer.

Immunoprecipitation. Wild-type or PINK1 knockout HeLa cells were transiently transfected with HA-tagged ubiquitin, ubiquitin(S65A) or ubiquitin(S65D) with or without mCherry-parkin for 24 h. Cells were collected, lysed and the HA-ubiquitin was immunoprecipitated as reported previously⁵, using anti-HA-conjugated beads (Pierce). To deubiquitinate the bound proteins, after binding the HA-ubiquitin, beads were washed three times and incubated in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM DTT and 1.47 μg USP2 (Boston Biochem) at 37 °C for 1 h. The reaction was stopped and the remaining bound proteins were washed five times with 1 ml of buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl), and then eluted by boiling with 1 \times LDS sample buffer.

In vitro phosphorylation. Strep-tagged ubiquitin was incubated with either wild-type or kinase-dead *Tribolium castaneum* PINK1 (TpPINK1) as previously reported⁵. This ubiquitin was then incubated with cytosol from wild-type HeLa cells in 20 mM HEPES-KOH, pH 7.6, 220 mM mannitol and 70 mM sucrose at 4 °C for 1 h. Strep-tactin beads (Qiagen) were then added to bind the strep-ubiquitin for an additional 1 h at 4 °C. The ubiquitin and bound proteins were then eluted with 50 mM biotin in 50 mM Tris for 15 min at room temperature. Samples were then diluted in LDS sample buffer before SDS-PAGE and immunoblot analysis.

Immunofluorescence microscopy. HeLa cells, seeded in two-well chamber slides (Lab-Tek), were treated as indicated in the figure legends. After treatment, cells were rinsed in PBS and fixed for 15 min at room temperature with 4% paraformaldehyde. Cells were then permeabilized and blocked with 0.1% Triton X-100, 3% goat serum in PBS for 40 min. For immunostaining, cells were incubated with antibodies (as indicated in figure legends) diluted in 3% goat blocking serum overnight at 4 °C, then rinsed with PBS and incubated with either anti-rabbit or mouse Alexa-Fluor-488- and Alexa-Fluor-633-conjugated secondary antibodies (Life Technologies), or anti-chicken Alexa-Fluor-488-conjugated antibody (Life Technologies) for 1 h at room temperature. Cells were washed three times for 5 min each with 1% Triton X-100 in PBS. During the final wash step, cells were incubated with 10 $\mu\text{g ml}^{-1}$ DAPI (Sigma) in PBS for 5 min. To measure mitophagy by mtDNA immunostaining, images were obtained from samples stained with DAPI and immunostained for DNA using a plan-Apochromat 63 \times /1.4 oil DIC objective on an LSM 510 microscope (Zeiss). Four image slices were collected through the Z plane encompassing the top and bottom of the cells. Image analysis was performed on all images collected in the Z plane using Volocity software (Perkin Elmer v6.0.1). The percentage of mtDNA stain remaining was calculated using the following formula: $(\text{cDNA}_v - \text{nDNA}_v)/n$, in which cDNA_v is the total cellular DNA volume determined by staining using anti-DNA antibodies, and nDNA_v is the total nuclear DNA stain volume determined using DAPI, and n denotes the number of cells.

The mtDNA stain volume in untreated cells was normalized to 100% and the amount of mtDNA stain remaining after drug treatment was subsequently determined. Final values represent data acquired from 50–200 cells from three independent experiments, and investigators were not blinded to allocation during image analysis.

To analyse LC3/mitochondrial protein colocalization; cells were treated, fixed and immunostained as above. Between 5 and 8 slices were imaged through the Z plane using either a plan-Apochromat 63× or 100×/1.4 oil DIC objective on a CW STED confocal microscope (Leica). Volocity software (Perkin Elmer, v6.0.1) was used to measure the intensity of the GFP signal representing LC3 in the volume occupied by mitochondria (as defined by the TOM20-positive region) and the cytosol (as defined by TOM20-negative region). 'Normalized mitochondrial LC3' was calculated using the following formula: normalized mitochondrial LC3 = $(m_i/m_v)/(c_i/c_v)$, in which m_i denotes the mitochondrial GFP intensity, m_v denotes the mitochondrial volume, c_i denotes the cytosolic GFP intensity, and c_v denotes the cytosolic volume. The resulting normalized mitochondrial LC3 is equal to 1 if the intensity of GFP is equal per volume in the cytosolic and mitochondrial volumes (no translocation) and is above one if the mitochondrial intensity is higher per volume (translocation). Final values for normalized mitochondrial LC3 represents data acquired from 50 to 105 cells from three independent experiments.

For GFP-DFCP1, GFP-WIPI1 and GFP-ULK1 puncta analysis, cells were treated, prepared and imaged on the CW STED as above with the addition of immunostaining using either rabbit or chicken GFP antibodies to enhance the signal in the green channel. GFP-DFCP1 puncta were quantified using Volocity software (Perkin Elmer v6.0.1) and GFP-WIPI1 and GFP-ULK1 puncta were quantified manually. Investigators were not blinded to allocation during image analysis. Colocalization of autophagy receptors with GFP-DFCP1 or GFP-ULK1 was assessed with line scans using LAS AF software (Leica, v2.6.0.7266).

Heterodimerization. The C-terminal Fis1 tail of human Fis1 (amino acids 92–152) was cloned into pC4-Rhe vector (ARIAD) at SpeI/BamHI sites to make FRB-Fis1 construct, the insert of which was then PCR amplified and cloned into pBMN-Z vector together with Puro-P2A sequence at HindIII, XhoI and NotI sites by an In-Fusion kit from Clontech to make pBMN-puro-P2A-FRB-Fis1. For receptor translocation assays, wild-type HeLa cells stably expressing FRB-Fis1 were generated using retroviral transduction as described above. Previously generated PINK1Δ110-YFP-2×FKBP (ref. 41) wild-type and kinase-dead and individually each mCherry-tagged autophagy receptor were transfected into FRB-Fis1 stable HeLa cells for 24 h. Cells were then treated with 0.5 μM rapalog (Clontech) for 8 h as previously described⁴¹. Cells were fixed and stained as described above. Cells were manually counted in a blinded manner for translocation of mCherry-tagged autophagy receptors to mitochondria. Final values represent data collected from 100–150 cells for three independent experiments.

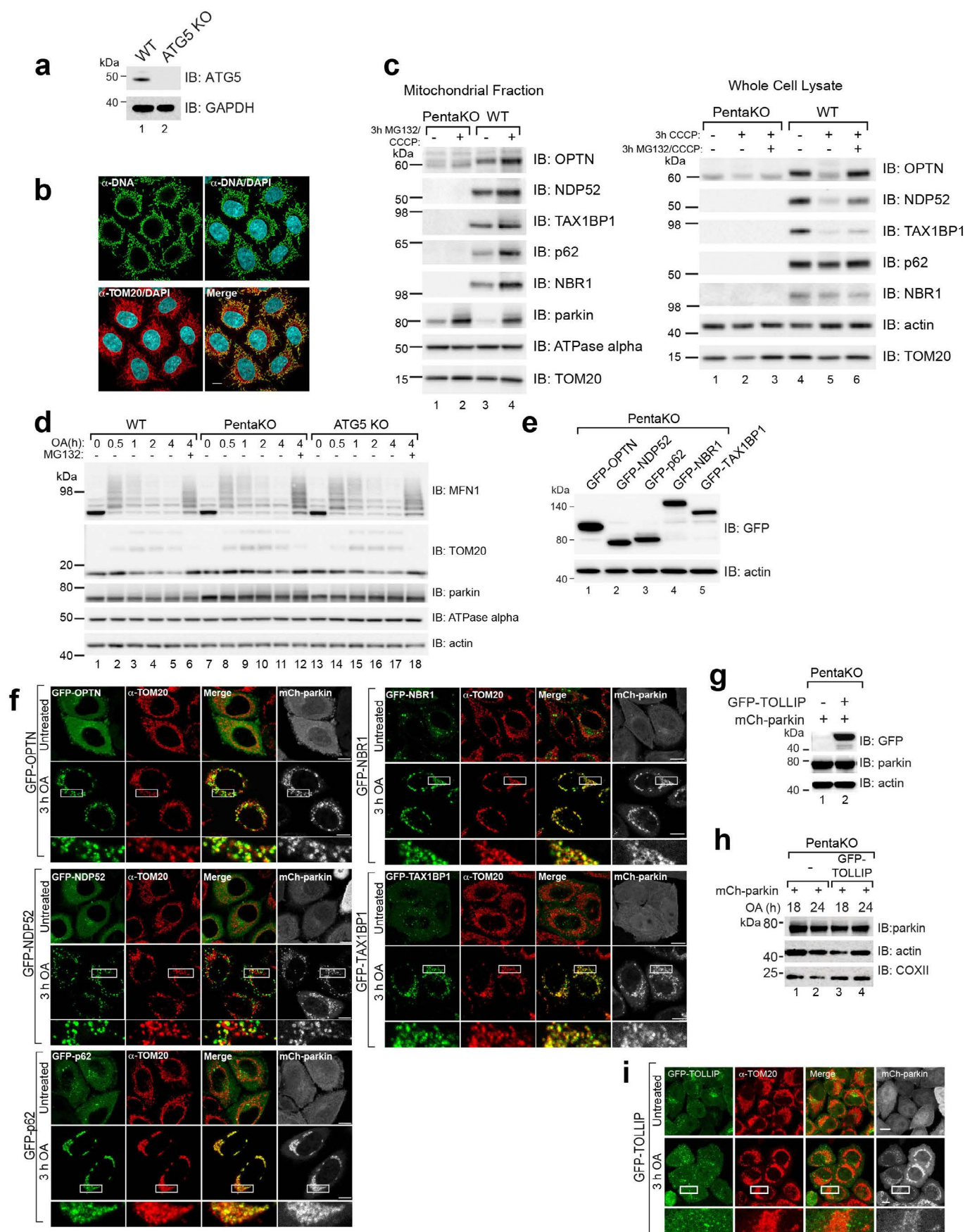
For ULK1 and DFCP1 rescue analysis, pentaKO HeLa cells stably expressing FRB-Fis1 were transiently transfected with PINK1Δ110-YFP-2×FKBP, mCherry-ULK1 and one of the autophagy receptors: Myc-OPTN, Myc-NDP52, Myc-OPTN(F178A), Myc-OPTN(E50K) or Myc-OPTN(Q398X) for 18–24 h. Cells were treated with 0.5 μM rapalog or vehicle for 7 h and imaged live on a Zeiss 780 in a humidified 37 °C, 5% CO₂ chamber. To visualize mitochondria in vehicle-treated controls, cells were pre-incubated for 10 min in 75 nM

Mitotracker Deep Red (Invitrogen) before imaging. Fields of PINK1-YFP-positive cells were imaged blindly to the mCherry-containing channel. The images were then blinded and counted manually for translocation of mCherry-tagged ULK1 to mitochondria. Final values represent >75 cells counted over at least two independent experiments.

Mito-Keima mitophagy assay. mt-mKeima⁴² (a gift from A. Miyawaki) was cloned into a pCHAC-MCS-1-IRES-MCS2 vector (Allele Biotechnology). PINK1Δ110-YFP-2×FKBP wild-type and kinase-dead were PCR-amplified from the original pC4M-F2E and cloned into pRetroQ-AcGFP-C1 at NheI/XhoI sites by Gibson assembly kit. The HA-tag was removed, and stop codon introduced. Wild-type and pentaKO HeLa cells stably expressing FRB-Fis1, mt-mKeima and either wild-type or kinase-dead PINK1Δ110-YFP-2×FKBP were generated using retroviral transduction as described above. Flag/HA-receptors were stably expressed in these cells by lentivirus transduction as described above then treated with 0.5 μM rapalog for 24 h. Cells were then resuspended in sorting buffer (145 mM NaCl, 5 mM KCl, 1.8 mM CaCl₂, 0.8 mM MgCl₂, 10 mM HEPES, 10 mM glucose, 0.1% BSA) containing 10 μg ml⁻¹ DAPI. Analysis was performed using Summit software (v6.2.6.16198) on a Beckman Coulter MoFlo Astrios cell sorter. Measurements of lysosomal mt-mKeima were made using dual-excitation ratiometric pH measurements at 488 (pH 7) and 561 (pH 4) nm lasers with 620/29 nm and 614/20 nm emission filters, respectively. For each sample, 50,000 events were collected and subsequently gated for YFP/mt-mKeima double-positive cells that were DAPI-negative. Data were analysed using FlowJo (v10, Tree Star).

Statistical calculations. No statistical methods were used to predetermine sample size, and experiments were not randomized. All statistical data were calculated and graphed using GraphPad Prism 6. To assess statistical significance, data from three or more independent experiments were analysed using one-way ANOVA and Tukey's post-test with a confidence interval of 95%. All error bars are expressed as mean ± s.d. In Fig. 4h, i, outliers were removed using ROUT in GraphPad Prism 6 with $Q = 1\%$, and 1–2 values from each condition were removed.

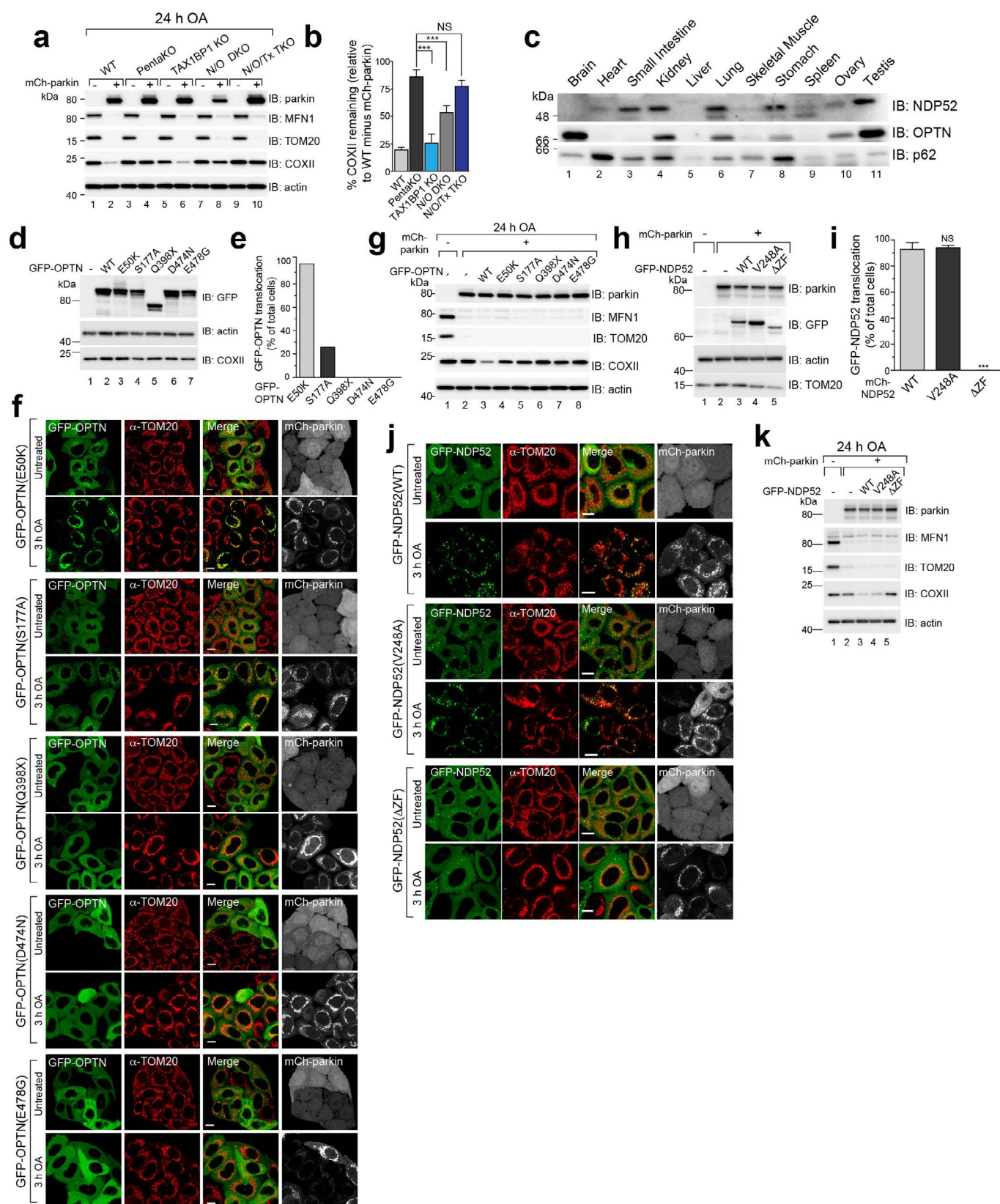
35. Nezhich, C. L., Wang, C., Fogel, A. I. & Youle, R. J. Mit/TFE transcription factors are activated during mitophagy downstream of Parkin and Atg5. *J. Cell Biol.* **210**, 435–450 (2015).
36. Santel, A. *et al.* Mitofusin-1 protein is a generally expressed mediator of mitochondrial fusion in mammalian cells. *J. Cell Sci.* **116**, 2763–2774 (2003).
37. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nature Biotechnol.* **29**, 699–700 (2011).
38. Hasson, S. A. *et al.* High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291–295 (2013).
39. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature Biotechnol.* **29**, 143–148 (2011).
40. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
41. Lazarou, M., Jin, S. M., Kane, L. A. & Youle, R. J. Role of PINK1 binding to the TOM complex and alternate intracellular membranes in recruitment and activation of the E3 ligase Parkin. *Dev. Cell* **22**, 320–333 (2012).
42. Katayama, H., Kogure, T., Mizushima, N., Yoshimori, T. & Miyawaki, A. A sensitive and quantitative technique for detecting autophagic events based on lysosomal delivery. *Chem. Biol.* **18**, 1042–1052 (2011).



Extended Data Figure 1 | Analysis of knockout cell lines and characterization of autophagy receptor translocation to damaged mitochondria.

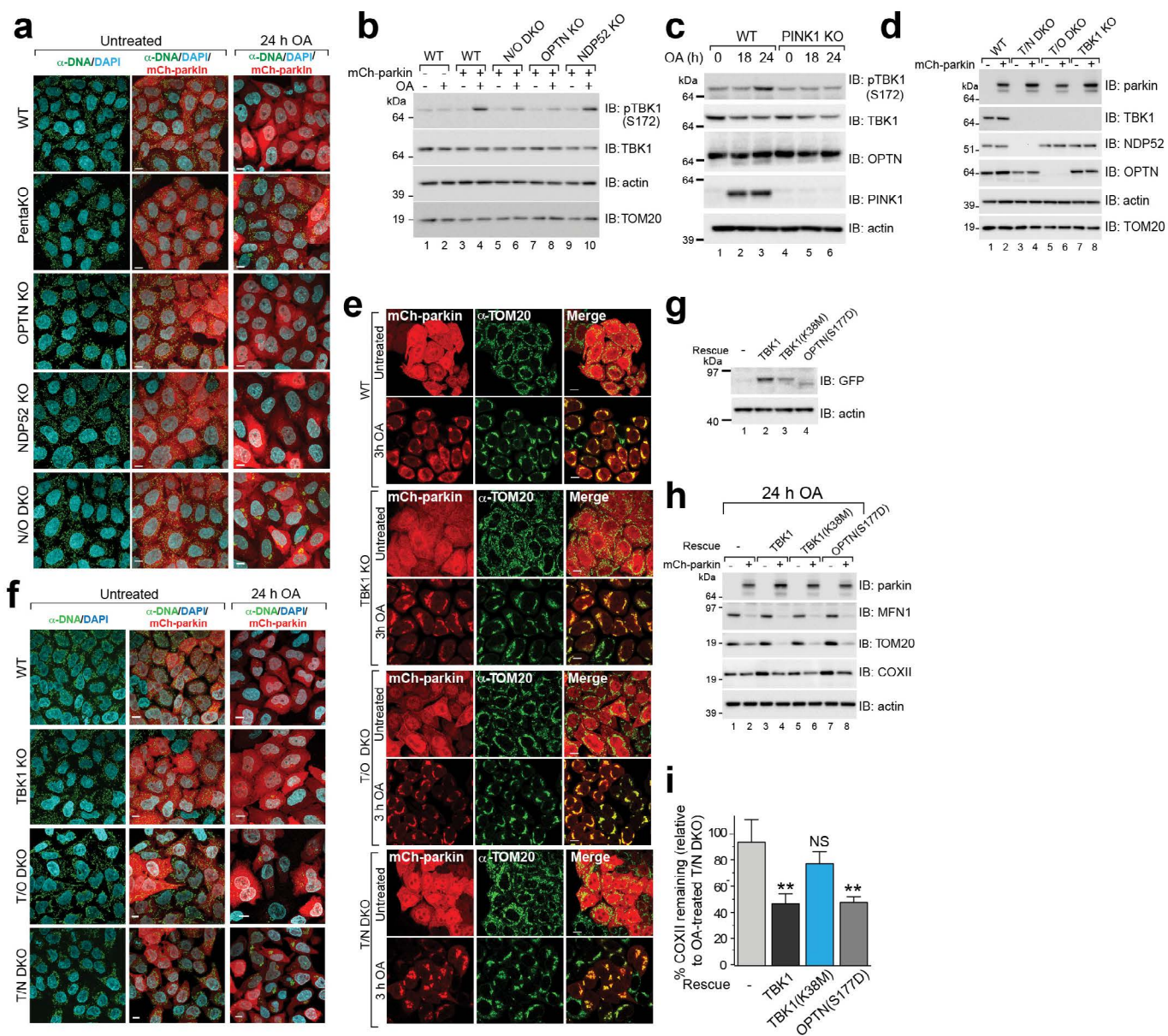
a, The ATG5 knockout cell line was confirmed by immunoblotting. **b**, Representative images of mtDNA nucleoids in HeLa cells immunostained with an anti-DNA antibody (green), confirming colocalization with the mitochondrial marker TOM20 (red) ($n = 3$ experiments). **c**, Mitochondrial fractions from mCherry-parkin expressing pentaKO and wild-type cells were assessed by immunoblotting. **d**, mCh-parkin-expressing wild-type, pentaKO and ATG5 knockouts were treated with oligomycin/

antimycin A with or without MG132. Cell lysates were assessed by immunoblotting. **e**, Expression levels of GFP-tagged OPTN, NDP52, p62, NBR1 and TAX1BP1 re-expressed in pentaKO cells by immunoblotting. **f**, Representative images of mCh-parkin-expressing pentaKOs from **e** immunostained for TOM20 ($n = 3$ experiments). **g**, Expression of GFP-TOLLIP in mCh-parkin pentaKOs. **h**, PentaKOs mCh-parkin and with or without GFP-TOLLIP expression were immunoblotted. **i**, Representative images of mCh-parkin pentaKOs expressing GFP-TOLLIP immunostained for TOM20 ($n = 3$ experiments). Scale bars, 10 μm .



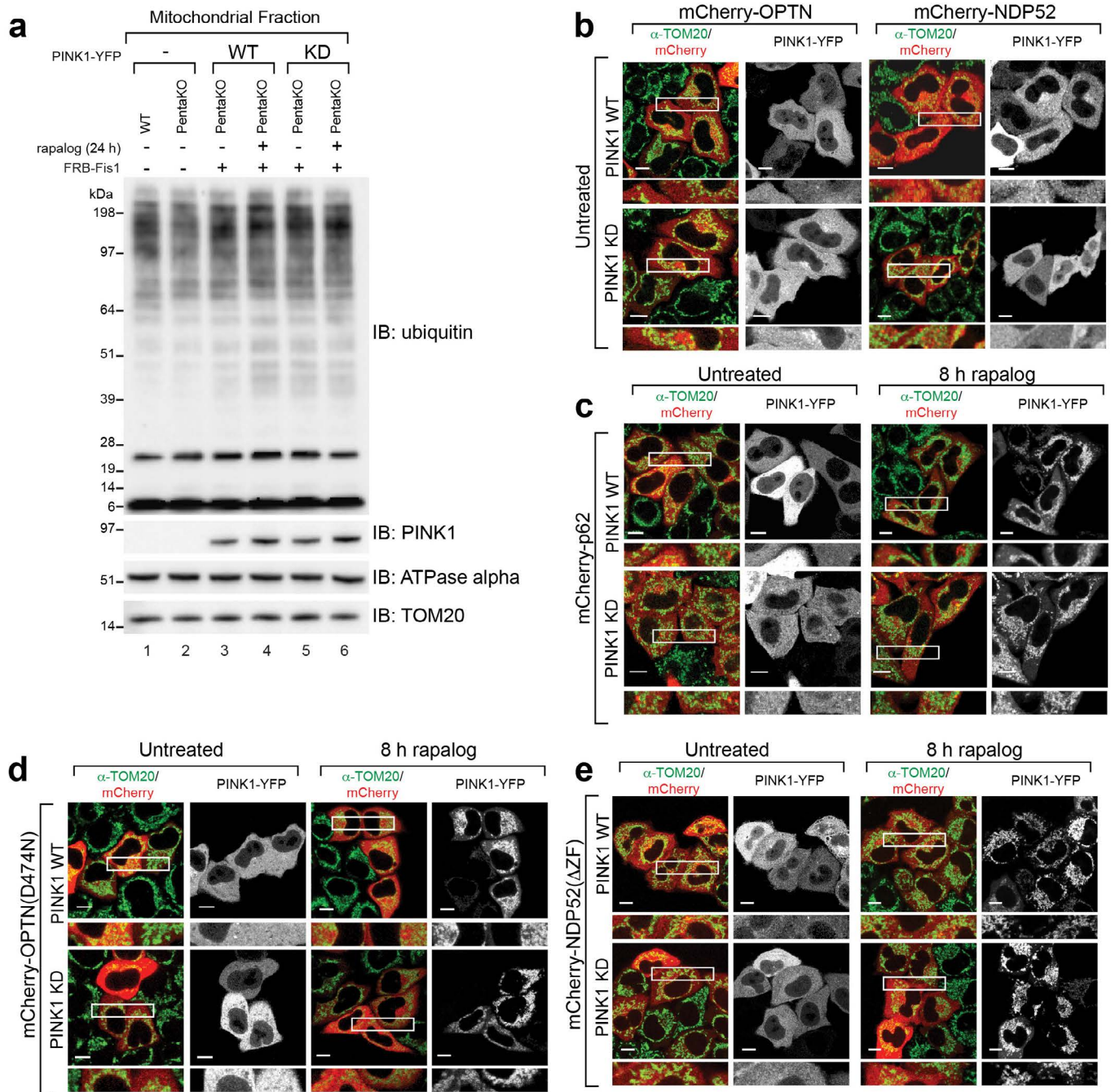
Extended Data Figure 2 | OPTN, NDP52 and TAX1BP1 triple knockout analysis and disease-associated mutations. **a, b**, Knockout cell lines with or without mCh-parkin expression were immunoblotted (**a**) and COXII levels were quantified (**b**). **c**, A panel of human tissue lysates was immunoblotted. **d**, Expression of wild-type or mutant GFP-OPTN in mCh-parkin pentaKO cells. **e**, Quantification of cells in **f**. More than 100 cells per condition. **f**, Representative images of mCh-parkin pentaKO cells expressing GFP-OPTN mutants immunostained for TOM20 ($n = 3$). **g**, PentaKOs expressing mCh-parkin were rescued with wild-type or mutant GFP-OPTN, analysed by

immunoblotting. See Fig. 2e for quantification of COXII. **h**, Expression of wild-type or mutant GFP-NDP52 in mCh-parkin pentaKO cells. **i**, Quantification of cells in **j**. More than 100 cells per condition. **j**, Representative images of mCh-parkin pentaKOs expressing wild-type or mutant GFP-NDP52 were immunostained for TOM20 ($n = 3$). **k**, PentaKO cells expressing mCh-parkin rescued with wild-type or mutant GFP-NDP52 were analysed by immunoblotting. See Fig. 2f for quantification of COXII. Data are mean \pm s.d. from three (**b** and **i**) and two (**e**) independent experiments. *** $P < 0.001$ (one-way ANOVA). Scale bars, 10 μ m.



Extended Data Figure 3 | TBK1 activates OPTN in PINK1/parkin mitophagy. **a**, Representative images of untreated mCh-parkin cells and merged images of treated cells as indicated immunostained for DNA. See Fig. 2a for anti-DNA and DAPI images of treated samples ($n = 3$). **b**, Cell lysates from wild-type, NDP52/OPTN DKO, OPTN knockout and NDP52 knockout cells with or without mCh-parkin expression were immunoblotted for TBK1 activation. **c**, Cell lysates from wild-type and PINK1 knockout cells without parkin expression were immunoblotted for TBK1 activation (S172 phosphorylation). **d**, Confirmation of TBK1/NDP52 DKO, TBK1/OPTN DKO

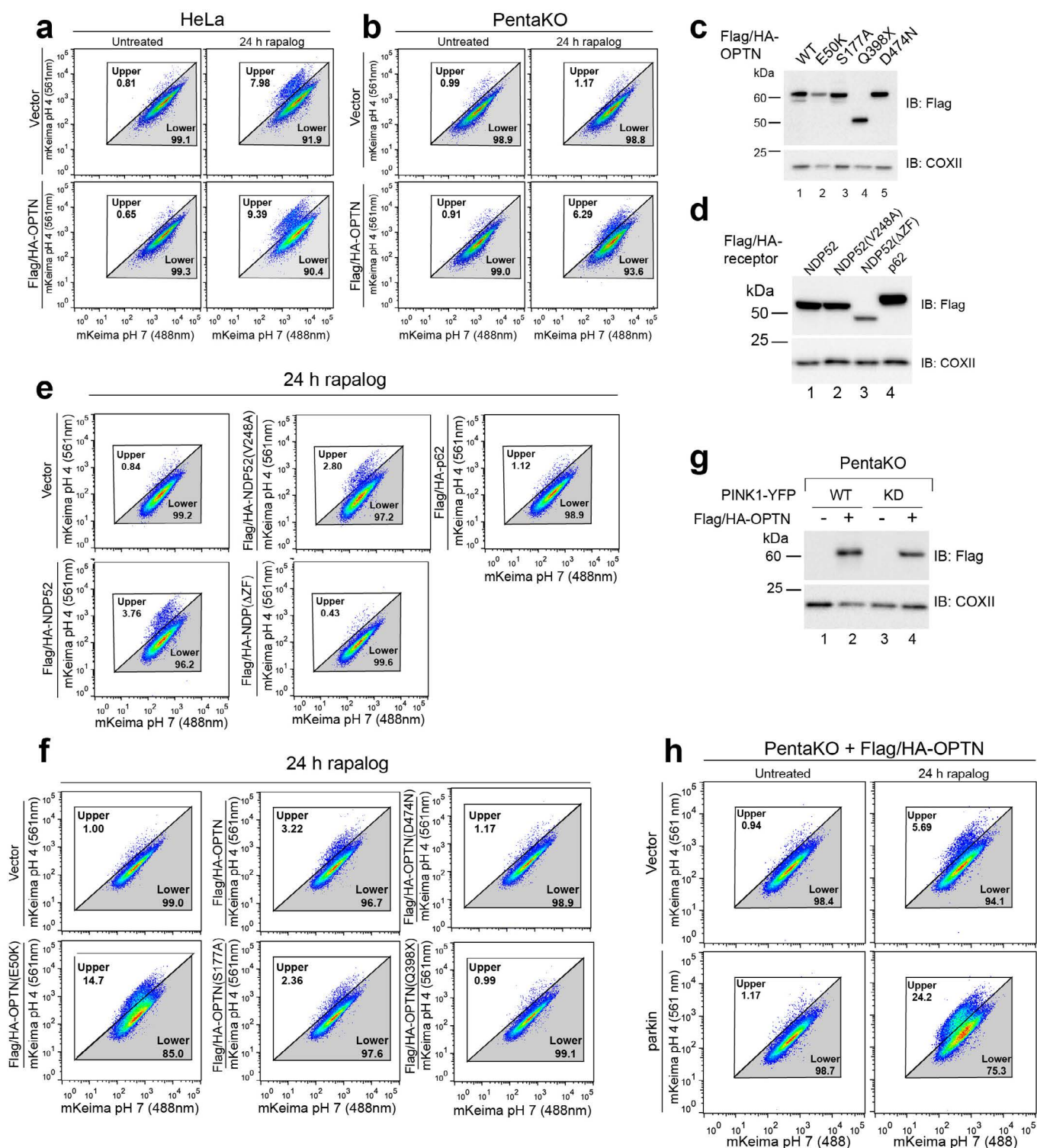
and TBK1 knockout by immunoblotting. **e**, Knockout cell lines from **d** were immunostained for TOM20 ($n = 3$). **f**, Representative images of untreated mCh-parkin wild-type and knockout cells, and merged images of treated cells as indicated were immunostained for DNA. See Fig. 2g for anti-DNA/DAPI images of treated samples ($n = 3$). **g**, TBK1/NDP52 DKO cells rescued with GFP-TBK1 wild-type or K38M, or GFP-OPTN(S177D) and were assessed by immunoblotting. **h**, Cells in **g** were assessed by immunoblotting. **i**, Quantification of COXII levels in **h** displayed as mean \pm s.d. from three independent experiments. ** P < 0.005 (one-way ANOVA). Scale bars, 10 μ m.



Extended Data Figure 4 | Parkin-independent recruitment of receptors to mitochondria through PINK1 activity. **a**, Isolated mitochondria from wild-type and pentaKO cells with or without FRB-Fis1 and with wild-type or kinase-dead PINK1 Δ 110-YFP-2 \times FKBP were immunoblotted.

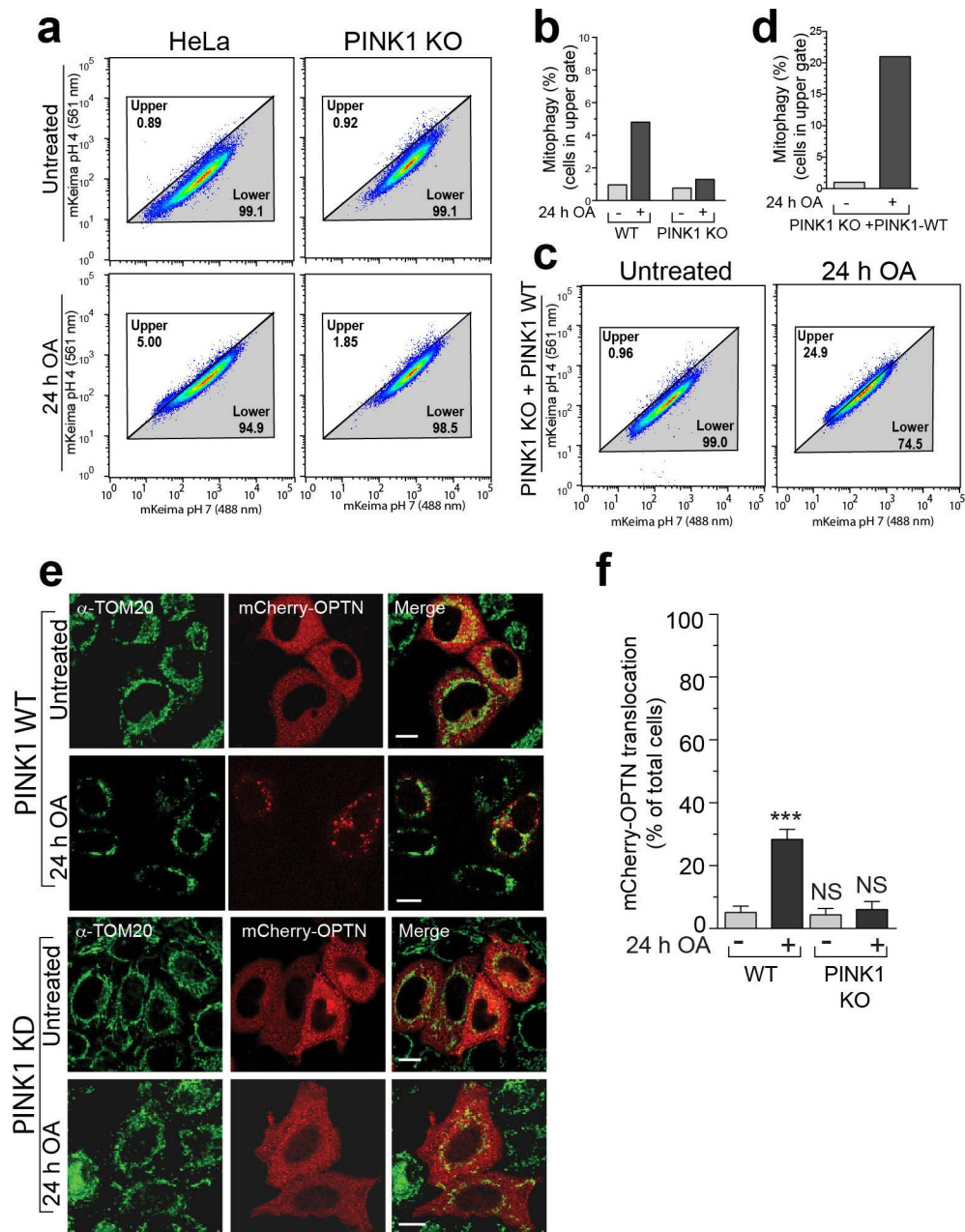
b–e, Representative images of pentaKO cells expressing FRB-Fis1, wild-type or

kinase-dead PINK1 Δ 110-YFP-2 \times FKBP, and mCh-OPTN or mCh-NDP52 (**b**), mCh-p62 (**c**), mCh-OPTN(D474N) (**d**) or mCh-NDP52(Δ ZF) (**e**). Cells were untreated (**b**) or treated with rapalog (**c–e**) and immunostained for TOM20. All images are representative of three independent experiments. See Fig. 3b for quantification. Scale bars, 10 μ m.



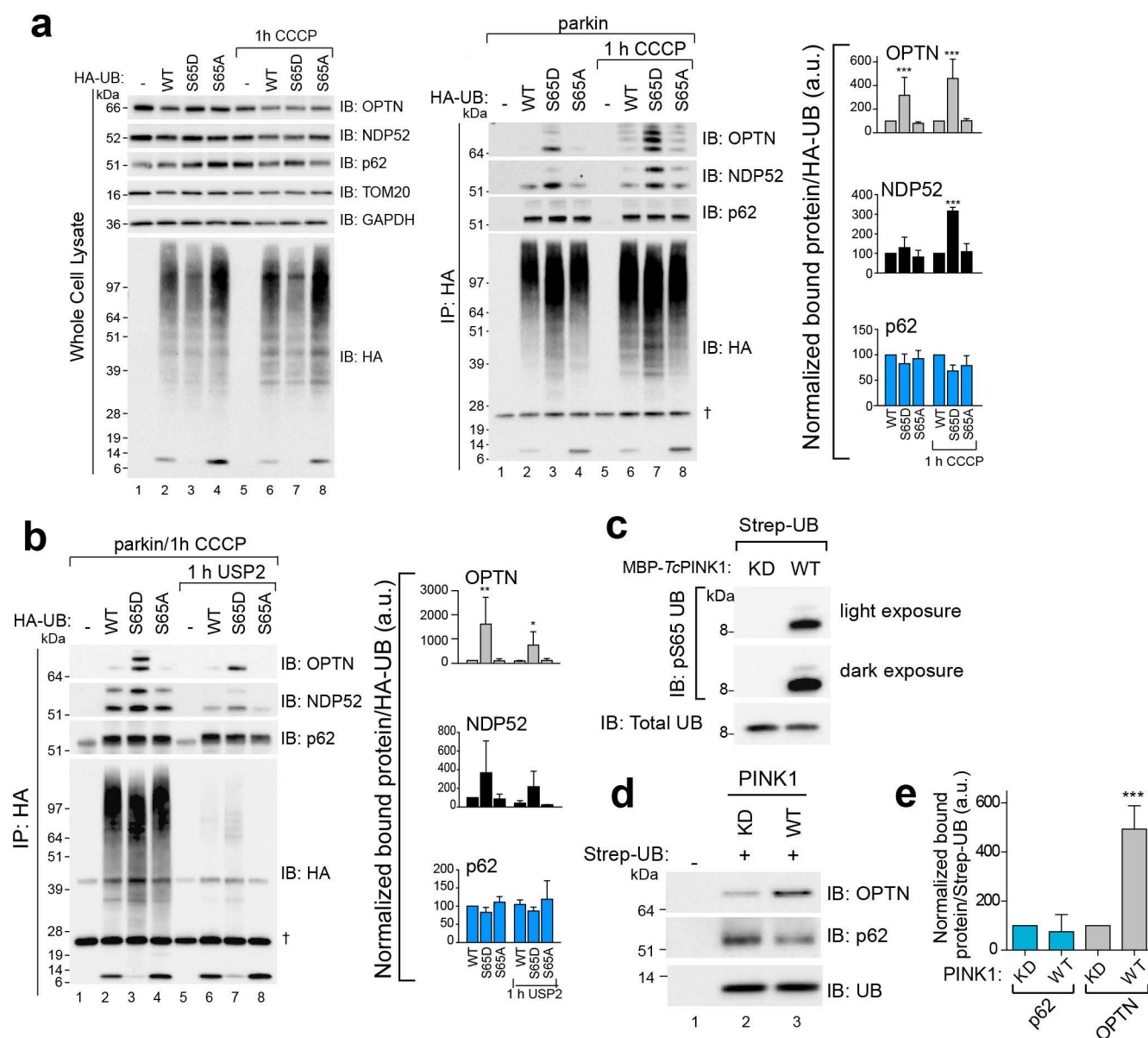
Extended Data Figure 5 | PINK1 directly stimulates mitophagy in the absence of mitochondrial damage. **a, b**, Cells were treated with rapalog and analysed by FACS for lysosomal-positive mt-mKeima. Representative data for wild-type HeLa (**a**) and pentaKO cells (**b**) without or with Flag/HA-OPTN. **c, d**, Cell lysates from pentaKO cells expressing FRB-Fis1, PINK1Δ110-YFP-2×FKBP, mt-mKeima and wild-type Flag/HA-OPTN (**c**) or mutants, (**d**) Flag/HA-p62, wild-type Flag/HA-NDP52 or NDP52 mutants as indicated were assessed for receptor expression by immunoblotting. **e, f**, Cells from **c** and

d were treated with rapalog and analysed by FACS for lysosomal-positive mt-mKeima. Representative data of two experiments are presented. **g**, Cell lysates from pentaKO cells expressing FRB-Fis1, with or without Flag/HA-OPTN and wild-type or kinase-dead PINK1Δ110-YFP-2×FKBP were assessed for OPTN by immunoblotting. **h**, Flag/HA-OPTN pentaKO cells expressing FRB-Fis1, PINK1Δ110-YFP-2×FKBP, mt-mKeima-transfected and either vector or untagged parkin were analysed by FACS. Representative data of two experiments are presented.



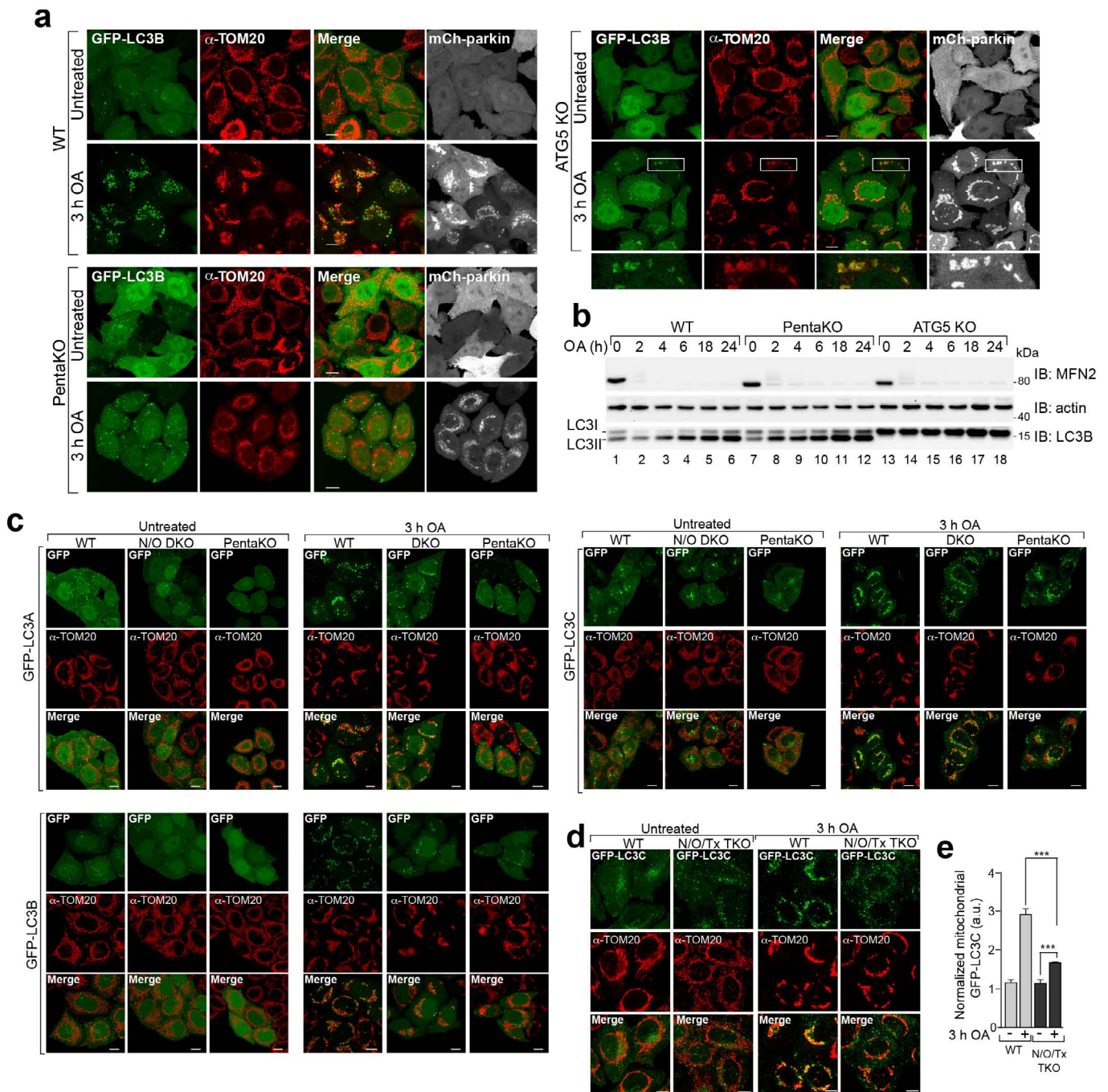
Extended Data Figure 6 | PINK1 directly stimulates mitophagy after mitochondrial damage. **a, c**, Representative FACS data of mt-mKeima-expressing wild-type or PINK1 knockout HeLa cells (**a**) or PINK1 knockout cells rescued with PINK1 wild-type and then untreated or treated with oligomycin and antimycin A (**c**). **b, d**, Average percentage of mitophagy for two replicates of **a** and **c**, respectively. **e**, Representative images of wild-type

HeLa cells expressing mCh-OPTN and treated with oligomycin and antimycin A as indicated and immunostained for TOM20 ($n = 3$). **f**, Quantification of mCh-OPTN translocation from cells in **e**. Data are mean \pm s.d. from three independent experiments. *** $P < 0.001$ (one-way ANOVA). NS, not significant.



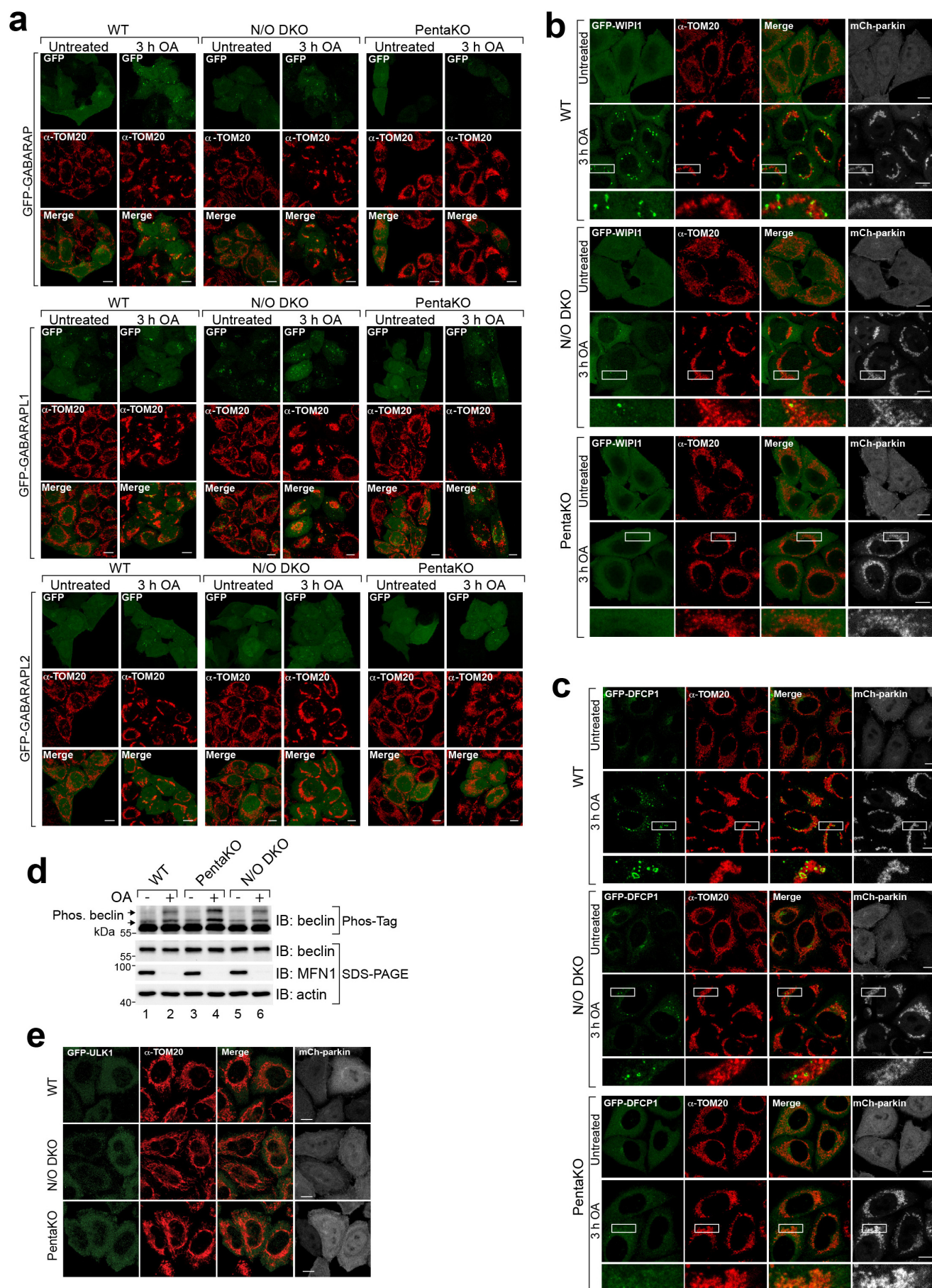
Extended Data Figure 7 | OPTN and NDP52 preferentially bind phosphomimetic ubiquitin. **a**, HeLa cells expressing mCh-parkin and wild-type HA-ubiquitin (HA-UB) or mutants S65D or S65A were treated with CCCP. HA-ubiquitin was co-immunoprecipitated and the bound fraction was analysed by immunoblotting. Quantifications of the total bound fraction of OPTN, NDP52 and p62 are shown. **b**, HA-ubiquitin transfected into HeLa cells with mCh-parkin were then treated with CCCP, and HA-ubiquitin was immunoprecipitated. The bound fraction was treated with the deubiquitinase USP2, and washed to remove all unbound protein after deubiquitination. Quantification of the total bound fraction of OPTN, NDP52 and p62 are shown

on the right. **c**, **d**, Strep-tactin-tagged ubiquitin (strep-UB) was incubated with either wild-type or kinase-dead PINK1 in an *in vitro* phosphorylation reaction, immunoblotted with an anti-phosphoS65 ubiquitin antibody (**c**), and then incubated with cytosol collected from untreated, wild-type HeLa cells. The ubiquitin was then pulled down using strep-tactin beads and analysed by immunoblotting (**d**). **e**, Quantification of bound OPTN and p62 normalized to total ubiquitin. Data in **a**, **b** and **e** are mean \pm s.d. from three independent experiments. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.001$ (one-way ANOVA). Dagger symbol denotes nonspecific band.



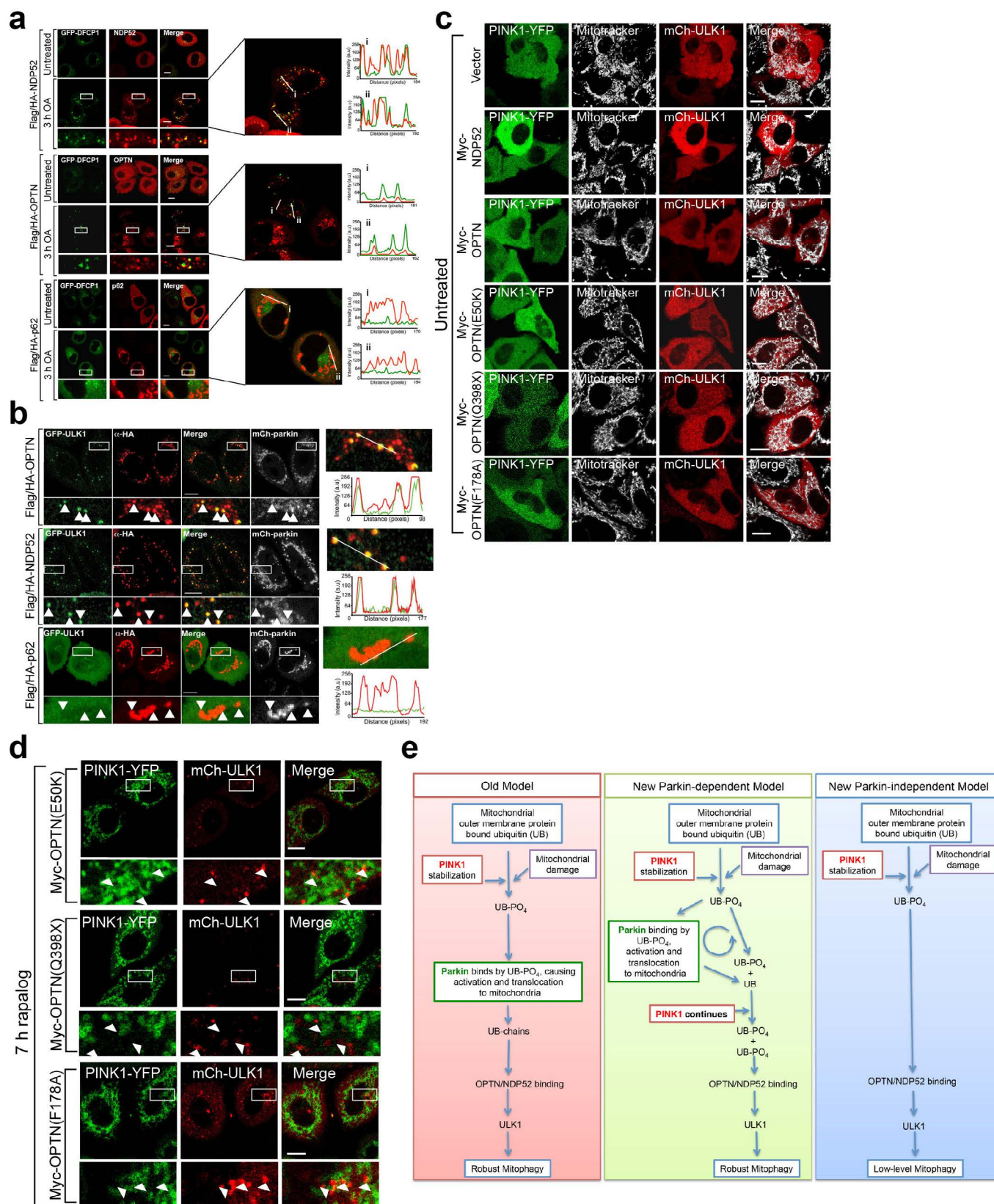
Extended Data Figure 8 | Analysis of LC3 family members and their translocation to damaged mitochondria in autophagy receptor knockout cell lines. **a**, Representative images of wild-type, pentaKO and ATG5 knockout HeLa cells expressing mCh-parkin and GFP-LC3B were immunostained for TOM20 ($n = 3$). **b**, Cell lysates from mCh-parkin expressing wild-type, pentaKO and ATG5 knockout cells were immunoblotted. **c**, Representative images of wild-type, NDP52/OPTN DKO and pentaKO cells expressing

mCh-parkin and GFP-tagged LC3A, LC3B or LC3C were immunostained for TOM20 ($n = 3$; see Fig. 4a for quantification). **d**, **e**, Representative images of wild-type and NDP52/OPTN/TAX1BP1 triple knockout cells expressing mCh-parkin and GFP-LC3C were immunostained for TOM20 ($n = 3$) (**d**) and quantified for GFP-LC3C translocation to mitochondria (**e**). Data in **e** are mean \pm s.d. from three independent experiments. *** $P < 0.001$ (one-way ANOVA). Scale bars, 10 μ m.



Extended Data Figure 9 | GABARAP proteins do not translocate to damaged mitochondria, and early stages of autophagosome biogenesis mediated by WIPI1 and DFCP1 are inhibited in autophagy-receptor-deficient cell lines. **a–c**, Representative images of wild-type, NDP52/OPTN DKO and pentaKO cells expressing mCh-parkin and GFP-tagged GABARAP, GABARAPL1 or GABARAPL2 (**a**), GFP-WIPI1 (**b**) or GFP-DFCP1 (**c**) immunostained for TOM20 ($n = 3$ for each condition, see Fig. 4b and c

for quantification of **b** and **c**). **d**, mCh-parkin cell lines as indicated were subjected to either phos-tag SDS-PAGE or standard SDS-PAGE followed by immunoblotting. Arrows indicate the position of phosphorylated beclin species. **e**, Representative images of untreated wild-type, NDP52/OPTN DKO and pentaKO cell lines expressing mCh-parkin and GFP-ULK1 were immunostained for TOM20 and GFP ($n = 3$ experiments). Scale bars, 10 μm .



Extended Data Figure 10 | OPTN and NDP52 rescue DFCP1 and ULK1 recruitment deficit in pentaKOs. **a**, Representative images of pentaKO cells expressing mCh-parkin, GFP-DFCP1 and the indicated Flag/HA-tagged autophagy receptors immunostained for haemagglutinin ($n = 2$ experiments). Panels on the right display co-localization of Flag/HA-tagged constructs and GFP-DFCP1 by fluorescence intensity line measurement. **b**, Representative images of pentaKO cells expressing mCh-parkin and GFP-ULK1 that were rescued by Flag/HA-OPTN, Flag/HA-NDP52 and Flag/HA-p62, and immunostained for haemagglutinin and GFP. Arrows indicate HA-tagged receptor puncta ($n = 2$). Right panels display colocalization of haemagglutinin and GFP by fluorescence intensity line measurement. **c**, **d**, Representative images of pentaKO cells stably expressing FRB-Fis1 and transiently expressing PINK1 Δ 110-YFP-2 \times FKBP and vector or Myc-tagged receptors, that were

untreated (**c**) or treated with rapalog (**d**) and imaged live ($n = 3$ experiments, see Fig. 4h, i for quantification of **c**, **d**). Scale bars, 10 μ m. **e**, Old and new models of PINK1/parkin mitophagy. The old model is dominated by parkin ubiquitination of mitochondrial proteins. In this, PINK1 has a small initiator role with the main function being to bring parkin to the mitochondria. The new model depicts parkin-dependent and -independent pathways leading to robust and low-level mitophagy, respectively. On the basis of our data, PINK1 is central to mitophagy both before and after parkin recruitment by phosphorylating ubiquitin to recruit parkin and autophagy receptors to mitochondria, to induce clearance. In the absence of parkin (right), this occurs at a low level owing to the relatively low basal ubiquitin levels on mitochondria. When parkin is present, it serves to amplify the PINK1 generated phospho-ubiquitin signal, allowing for robust and rapid mitophagy induction.

Structural insights into μ -opioid receptor activation

Weijiao Huang^{1*}, Aashish Manglik^{1*}, A. J. Venkatakrishnan^{1,2,3}, Toon Laeremans^{4,5}, Evan N. Feinberg^{1,2,3}, Adrian L. Sanborn^{1,2,3}, Hideaki E. Kato¹, Kathryn E. Livingston⁶, Thor S. Thorsen¹, Ralf C. Kling⁷, Sébastien Granier⁸, Peter Gmeiner⁷, Stephen M. Husbands⁹, John R. Traynor⁶, William I. Weis^{1,10}, Jan Steyaert^{4,5}, Ron O. Dror^{1,2,3} & Brian K. Kobilka¹

Activation of the μ -opioid receptor (μ OR) is responsible for the efficacy of the most effective analgesics. To shed light on the structural basis for μ OR activation, here we report a 2.1 Å X-ray crystal structure of the murine μ OR bound to the morphinan agonist BU72 and a G protein mimetic camelid antibody fragment. The BU72-stabilized changes in the μ OR binding pocket are subtle and differ from those observed for agonist-bound structures of the β_2 -adrenergic receptor (β_2 AR) and the M2 muscarinic receptor. Comparison with active β_2 AR reveals a common rearrangement in the packing of three conserved amino acids in the core of the μ OR, and molecular dynamics simulations illustrate how the ligand-binding pocket is conformationally linked to this conserved triad. Additionally, an extensive polar network between the ligand-binding pocket and the cytoplasmic domains appears to play a similar role in signal propagation for all three G-protein-coupled receptors.

The most powerful analgesic and addictive properties of opiate alkaloids are mediated by the μ OR¹. As the receptor primarily responsible for the effects of opium, the μ OR is one of the oldest drug targets within the pharmacopeia². Opioid receptors are highly versatile signalling molecules. Activation of the μ OR results in signalling through the heterotrimeric G protein G_i , resulting in analgesia and sedation as well as euphoria and physical dependence³. The μ OR can also signal through arrestin, and this pathway has been attributed to adverse effects of opioid analgesics including tolerance, respiratory suppression, and constipation^{4–6}.

The μ OR has been the subject of intense focus for drug-discovery efforts over the past century, with the identification of numerous ligands of varying efficacy. These drugs occupy a wide chemical spectrum, from small organic molecules to a variety of endogenous and synthetic peptides⁷. Structure-activity studies have revealed that subtle changes in ligand structure can convert an agonist into an antagonist⁷. These studies have yielded a general hypothesis for the information encoded within G-protein-coupled receptor (GPCR) ligands where distinct pharmacophores within a drug are responsible for efficacy (message) or selectivity (address)⁸ (Fig. 1a). For the morphinan ligands, our previous structural examination of the inactive states of the μ OR and the δ OR revealed molecular insights into ligand selectivity^{9,10}. To understand the structural basis for μ OR activation, we obtained a structure of this receptor in the active state using a combination of a high-affinity agonist and a G protein mimetic camelid antibody fragment. A comparison of this structure with the inactive-state structures of the μ OR⁹ and δ OR^{10,11}, as well as the inactive- and active-state structures of the β_2 AR^{12–15}, M2 muscarinic receptor (M2R)^{16,17}, and rhodopsin^{18,19}, provide insights into shared mechanisms of GPCR activation.

Nanobody-stabilized structure of the μ OR

The active states of ligand-activated GPCRs are probably unstable, even when bound to full agonists^{20–23}. However, the active conformation can be stabilized by interactions between a receptor and its cognate G protein. This stabilization is reflected in a higher affinity for agonists when GPCRs are in complex with their cognate G protein²⁴. In the case of the μ OR, the affinity for the morphinan agonist BU72 is enhanced by 47-fold when coupled to the G protein G_i (Fig. 1b, c). Efforts to obtain a structure of activated μ OR in complex with G_i have thus far not been successful. As an alternative, we have previously used camelid single-domain antibody fragments (nanobodies) as G protein mimetics to stabilize the active conformation of the β_2 AR and M2R for structural study^{12,13,17}. For the β_2 AR, the conformation of the receptor obtained in complex with the G_s mimetic nanobody 80 (Nb80) was nearly identical to that in the β_2 AR- G_s complex²⁵ (root mean squared deviation, 0.61 Å).

To generate G protein mimetic nanobodies for the μ OR, llamas were immunized with purified μ OR bound to the peptide agonist [Dmt¹]DALDA (ref. 26) and reconstituted into phospholipid vesicles¹². We examined the ability of selected nanobodies to stabilize the high-affinity state for μ OR agonists. Purified μ OR was reconstituted into high-density lipoprotein (HDL) particles and agonist competition assays were performed in the presence or absence of nanobodies (Fig. 1b). In the presence of 5 μ M nanobody 39 (Nb39), the affinity of the potent morphinan agonist BU72²⁷ increases from 470 pM to 16 pM (Fig. 1b). BU72 has a dissociation half-life of 140 min in the presence of Nb39 (Extended Data Fig. 1b). Nb39 also enhances the affinity of μ OR agonists DAMGO and endomorphin-2, indicating that the effect is not limited to morphinan agonists (Extended Data Fig. 1a).

¹Department of Molecular and Cellular Physiology, Stanford University School of Medicine, 279 Campus Drive, Stanford, California 94305, USA. ²Department of Computer Science, Stanford University, 318 Campus Drive, Stanford, California 94305, USA. ³Institute for Computational and Mathematical Engineering, Stanford University, 475 Via Ortega, Stanford, California 94305, USA. ⁴Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. ⁵Structural Biology Research Center, VIB, Pleinlaan 2, B-1050 Brussels, Belgium. ⁶Department of Pharmacology, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁷Department of Chemistry and Pharmacy, Friedrich Alexander University, Schuhstrasse 19, 91052 Erlangen, Germany. ⁸Institut de Génomique Fonctionnelle, CNRS UMR-5203 INSERM U1191, University of Montpellier, F-34000 Montpellier, France. ⁹Department of Pharmacy and Pharmacology, University of Bath, Bath BA2 7AY, UK. ¹⁰Department of Structural Biology, Stanford University School of Medicine, 299 Campus Drive, Stanford, California 94305, USA.

*These authors contributed equally to this work.

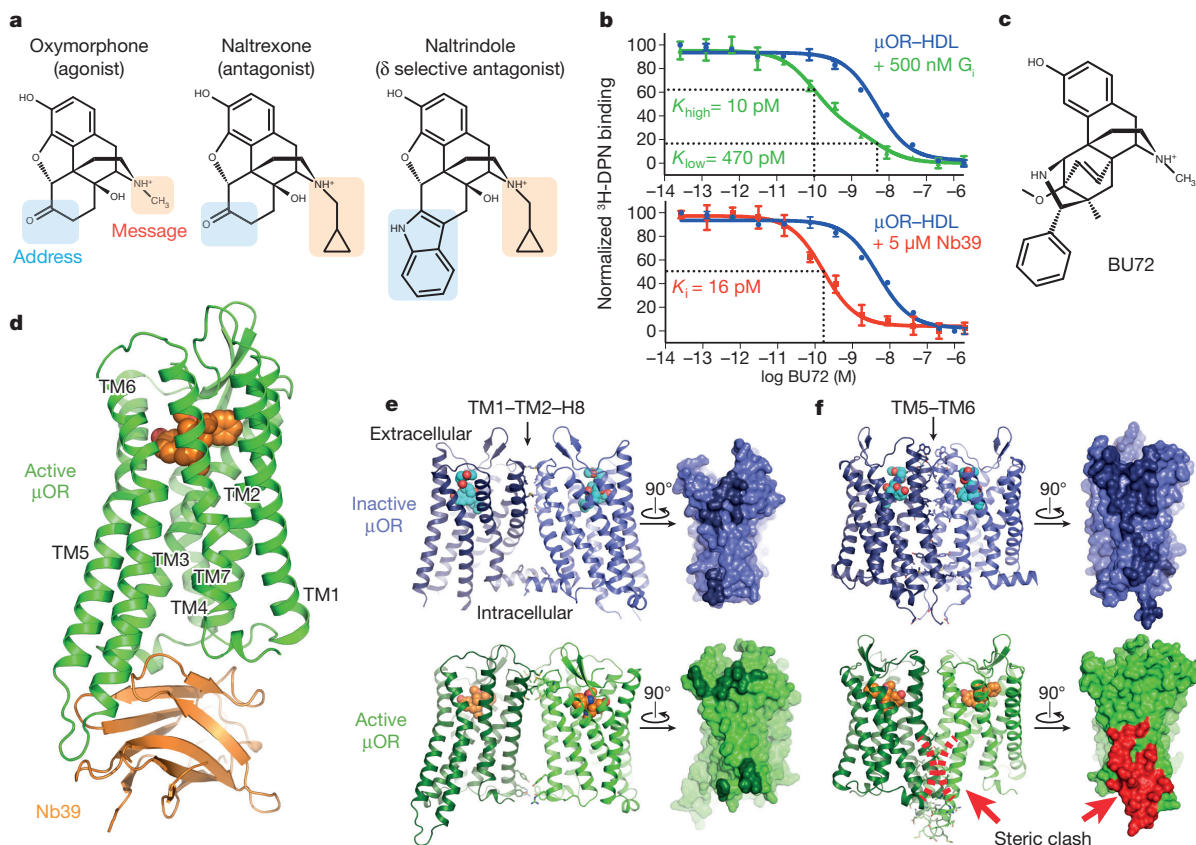


Figure 1 | Activated structure of μ OR bound to BU72 and Nb39.

a, Structures of prototypical opioid ligands highlighting regions involved in encoding efficacy (message) and selectivity (address). **b**, ^3H -diprenorphine (^3H -DPN) radioligand competition binding of μ OR in HDL particles. In the presence of G_i , the affinity of the morphinan agonist BU72 increases 47-fold. The two observed binding sites indicate the affinity of BU72 for receptor coupled to G_i (K_{high}) and uncoupled to G_i (K_{low}). A similar 29-fold increase in affinity is observed in presence of Nb39. The binding curves are representative

of at least three experiments performed in triplicate, and the data and error bars represent the mean \pm s.e.m. **c**, Structure of the high affinity agonist BU72. **d**, Overall structure of the μ OR-BU72-Nb39 complex. **e**, An interface between TM1-TM2 and helix 8 (H8) is observed in both inactive and active structures of the μ OR. The residues comprising the interface are highlighted in dark colours on the surface view. **f**, The TM5-TM6 interface observed for inactive μ OR is not compatible with the active state due to clashing residues in TM5 and TM6 (highlighted in red).

Crystals of the μ OR bound to BU72 and Nb39 were obtained in a monoolein lipidic mesophase²⁸ and a complete data set to 2.1 Å was obtained by merging diffraction data from four crystals (Extended Data Table 1). Nb39 binds to the intracellular surface of μ OR (Fig. 1d and Extended Data Fig. 2) and mediates the majority of packing interactions between lipidic layers and adjacent Nb39- μ OR complexes in the crystal lattice. There are no packing interactions involving the extracellular surface of the receptor (Extended Data Fig. 1c). We observe a limited parallel dimeric packing interaction between μ OR molecules involving the extracellular end of transmembrane helix 1 (TM1), TM2, the first extracellular loop (ECL1), and helix 8 with a buried surface area of 460 Å² (Fig. 1e). A similar interface between TM1, TM2, and helix 8 was also observed in the inactive structure of μ OR with a slightly larger buried surface area of 615 Å² (Fig. 1e). The inactive structure also identified a more extensive parallel dimer interaction involving TM5 and TM6 with a buried surface area of 1,460 Å² (Fig. 1f). This interaction involving TM5 and TM6 would not be compatible with the conformational changes we observe in the active state (Fig. 1f). It is important to note that the physiological relevance of these interfaces remains unclear.

Structural differences in the extracellular surface between inactive and active μ OR are relatively small (Fig. 2a) with the exception of the proximal N terminus, as discussed below. Conformational changes at the cytoplasmic surface of the μ OR observed upon activation are similar to those observed for the β_2 AR, M2R and rhodopsin, with

large outward movement of TM6 and a smaller inward movement of TM5 and TM7 (Fig. 2a and Extended Data Fig. 3). The conserved E/DRY motif at the intracellular end of TM3 plays a role in maintaining GPCRs in the inactive state. In rhodopsin, an ionic interaction between R135^{3,50} and E247^{6,30} in TM6 stabilizes TM6 in an inactive conformation (superscript numbers follow the Ballesteros-Weinstein numbering method for GPCRs²⁹). While there is no acidic amino acid at the end of TM6 of μ OR that can form a similar salt bridge with R165^{3,50}, R165^{3,50} can form a hydrogen bond with T279^{6,34} (Fig. 2b). In the active states of the μ OR and metarhodopsin II, R^{3,50} forms a hydrogen bond with Y^{5,58}, stabilizing the inward movement of TM5 (Fig. 2b). R^{3,50} and Y^{5,58} assume a similar orientation in the M2R and the β_2 AR- G_s complex; however, they are not close enough to form a hydrogen bond (Extended Data Fig. 3a).

Agonist binding pocket
The morphinan scaffold of BU72 binds to the activated μ OR in a similar orientation to that observed previously for the irreversible antagonist β -funaltrexamine (β -FNA) at the inactive μ OR (Fig. 3a and Extended Data Fig. 4). Fig. 3b shows that the overall structural differences in the orthosteric binding pockets of active and inactive μ OR are relatively subtle. The majority of interactions between BU72 and active μ OR are hydrophobic or aromatic in nature, with the exception of two conserved polar interactions (Fig. 3c and Extended Data Fig. 4). As observed previously in the inactive structures of the

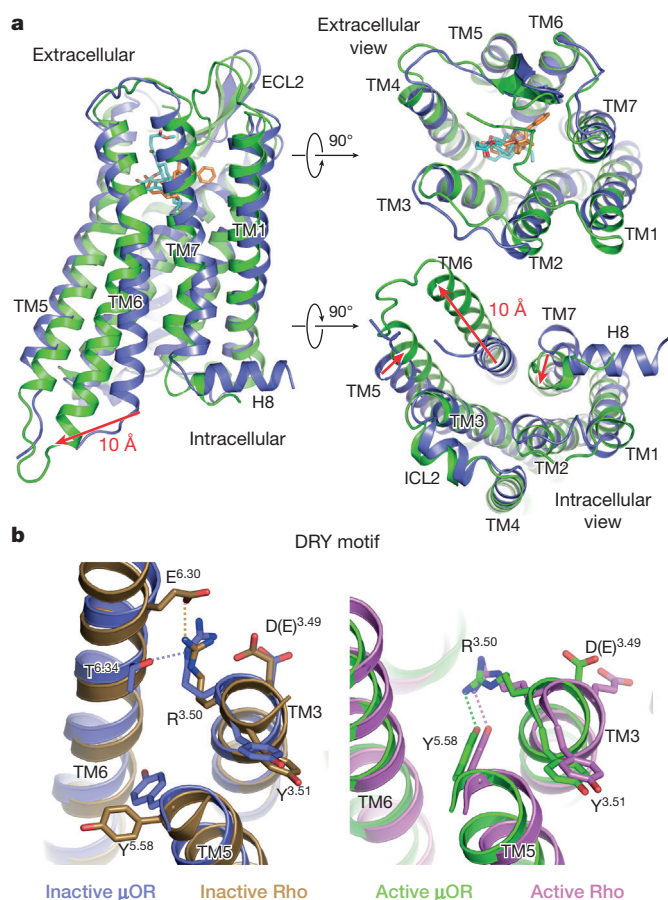


Figure 2 | Structural comparison of inactive and active μ OR. **a**, Active μ OR undergoes a 10 Å outward displacement of TM6 on activation. The extracellular domain of the receptor shows minimal changes upon activation. H8, helix 8. **b**, Left, comparison of the conserved E/DRY motif in the inactive structures of μ OR (blue) and rhodopsin (Rho, brown) shows a polar interaction between R^{3.50} and T^{6.34} in μ OR, analogous to the ionic lock between R^{3.50} and E^{6.30} observed for rhodopsin. Right, comparison of the same region in the active state of μ OR (green) and Rho (purple) shows a conserved interaction between R^{3.50} and Y^{5.58}.

μ OR, δ OR and κ OR³⁰, the phenolic hydroxyl of BU72 engages in a water-mediated interaction with H297^{6.52}. In the active state, this network is more extended, and involves Y148^{3.33} and the backbone carbonyl of K233^{5.39}. An ionic interaction between the morphinan tertiary amine and D147^{3.32} is seen in both the active μ OR bound to BU72 and in inactive μ OR bound to β -FNA (Fig. 3c).

We also observe an unexpected interaction between BU72 and the amino terminus of the μ OR, which forms a lid over the ligand-binding pocket (Fig. 3d and Extended Data Fig. 5). In particular, the amino-terminal residue H54 is positioned 2.6 Å from secondary amine of BU72. It is of interest that binding of BU72, as well as other structurally unrelated peptide agonists like DAMGO and [Dmt¹]DALDA, lead to conformational changes in the truncated amino terminus that can be detected by nuclear magnetic resonance (NMR) spectroscopy (Sounier *et al.*, ref. 31). While interactions between the amino terminus and the transmembrane core may indeed be important for μ OR function³², the specific interaction observed here is unlikely to be physiologically relevant as H54 is not highly conserved and the H54A mutation does not alter the affinity of BU72 in μ OR with the full-length amino terminus (μ OR wild-type K_i = 21 pM and μ OR H54A K_i = 30 pM in HEK293 membranes). Additionally, using a bioluminescence resonance energy transfer assay³³, we observe no significant difference in the EC₅₀ of G_i activation by BU72 between

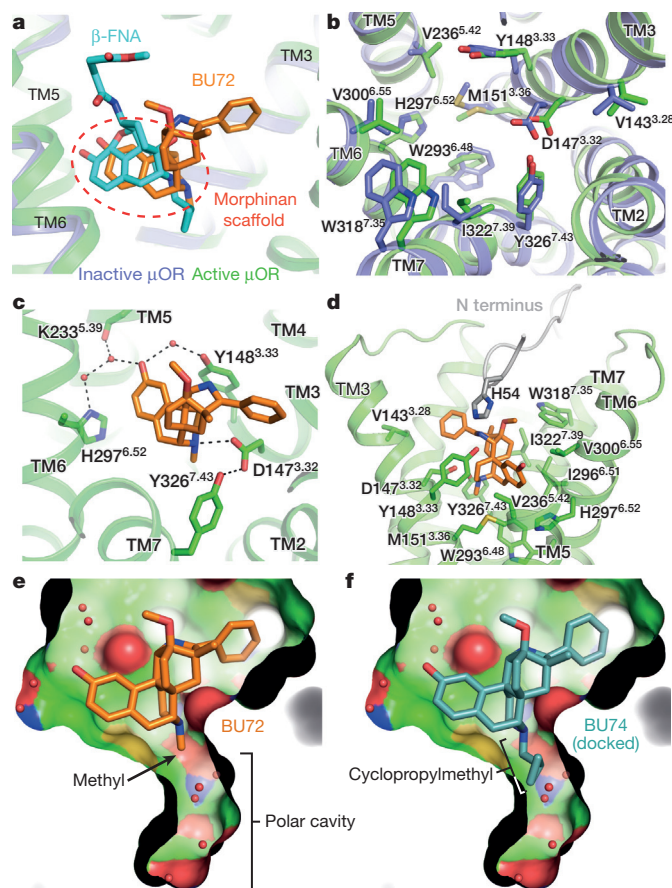


Figure 3 | The μ OR agonist-binding pocket. **a**, BU72 and β -funaltrexamine (β -FNA) occupy a similar pose in the μ OR binding pocket. The common morphinan scaffold shared by both ligands is highlighted. **b**, Binding-pocket residues of inactive (blue) and active (green) μ OR viewed from the extracellular side. **c**, Polar interactions between BU72 and active μ OR. **d**, BU72 and binding-pocket residues shown with the proximal amino terminus (grey). **e**, View of the polar cavity extending towards the intracellular side of μ OR in the active state. **f**, The cyclopropylmethyl group of the antagonist BU74 can be docked to fit within the polar cavity of the active state. However, molecular dynamics simulations show that this pose is unstable (Extended Data Fig. 6).

the wild-type receptor and the H54A mutant (EC₅₀ for wild type of 79 ± 17 pM and H54A mutant of 67 ± 20 pM).

While the morphinan core of BU72 is unambiguously placed within the electron density, there are two unexplained features of the ligand. We observe a strong positive electron density ($F_o - F_c$ signal) between the H54 side chain and the BU72 amine (Extended Data Fig. 5c). Attempts to identify the source of this density, including mass spectrometry for an alternative ligand structure and the presence of an anomalous signal for a coordinated metal were unsuccessful (Extended Data Fig. 4). Another unexpected finding is the near-planar geometry of the sp^3 hybridized carbon of BU72 to which the pendant phenyl group is attached (Extended Data Fig. 4e–h). This geometry could be explained by a double bond between this carbon and the adjacent nitrogen. We observed a minor fraction of such a compound by mass spectrometry in our preparation of BU72. However, this compound was not observed in mass spectrometry of the μ OR–ligand–Nb39 complex used for crystallography. Ultimately, we modelled a higher-energy conformation of BU72 within the observed electron density (see Extended Data Fig. 4 for further discussion).

In the active μ OR, there is a water-filled cavity lined by polar and aromatic side chains that extends off of the intracellular end of the ligand-binding pocket (Fig. 3e). While a similar cavity is observed in

inactive μ OR, the active-state cavity is larger and completely contiguous with the morphinan-binding site. Substitution of morphinans with a cyclopropylmethyl at the tertiary amine generally results in a ligand with antagonist activity⁷. As an example, BU74, which only differs from BU72 in having a cyclopropylmethyl substituent on the tertiary amine (Fig. 3f), is an antagonist at the μ OR³⁴. While BU74 can be docked into the active state μ OR structure with the cyclopropylmethyl within the polar cavity, there are potential clashes with Y326^{7,43} in TM7 and W293^{6,48} in TM6. Moreover, the cyclopropylmethyl substituent would displace one or more water molecules that form part of a polar network described in more detail below. Consistent with these observations, molecular dynamics simulations reveal that the antagonist BU74 is unstable in the position occupied by the agonist BU72 and rapidly shifts away from this initial pose (Extended Data Fig. 6).

Propagation of conformational changes

The structural difference in the orthosteric ligand-binding pocket between active and inactive μ OR is relatively subtle (Fig. 3b). It is difficult to identify specific interactions between the structurally rigid agonist and the receptor as being key for μ OR activation. This stands in contrast to activation of the β_2 AR and M2R, where specific polar interactions between the receptor and the smaller, more flexible agonists contribute to structural changes associated with activation. These polar interactions stabilize a 2 Å inward movement of TM5 in the β_2 AR and a 2 Å inward movement of TM6 in the M2R.

While agonist-stabilized changes in the binding pocket differ for the μ OR, β_2 AR, and M2R, the overall structural changes observed at the G-protein interface are very similar (Extended Data Fig. 3). The mechanism by which agonist-stabilized changes propagate to the cytoplasmic surface appears to be more similar between μ OR and the β_2 AR compared to the M2R (Fig. 4a). For both the β_2 AR and

the μ OR, there is a rearrangement of the packing of a triad of conserved amino acids F^{6,44}, P^{5,50} and I^{3,40} (which we term the conserved core triad) that lay just below the binding pocket (Fig. 4a). This rearrangement is associated with a counter-clockwise rotation (when viewed from the extracellular surface) and outward movement of the cytoplasmic end of TM6. Thus, the subtle agonist-stabilized rearrangement in the conserved core triad may initiate the cascade of structural changes involved in activation of the μ OR and β_2 AR. The role of the conserved core triad appears to differ in the M2R, where the smaller V111^{3,40} forms a weaker packing interaction with F396^{6,44} and P198^{5,50}. As a result, the triad in both inactive and active states of the M2R appear similar to the active states of the β_2 AR (Fig. 4a).

Given the similarity in allosteric propagation between μ OR and β_2 AR, we sought to identify how BU72 stabilizes the active conformation of the conserved core triad. In the β_2 AR, the rearrangement of the conserved core triad can be attributed to an agonist-stabilized inward movement of TM5^{12,13}. For the μ OR, however, there are no specific interactions between the receptor and BU72 that stabilize the inward movement of TM5. Instead, we identify a set of interactions that together appear to stabilize the rearrangement of the conserved core triad in the μ OR.

Both β -FNA and BU72 share a common morphinan scaffold. This common scaffold, however, is positioned differently in the inactive and active structures. In both states, residues I296^{6,51} and V300^{6,55} in TM6 and W318^{7,35} and I322^{7,39} in TM7 form a common hydrophobic surface for binding the morphinan ligands (Fig. 4b, d, spheres). While this surface is similar in both inactive and active states, differences in the position of the rigid morphinan agonist and antagonist result in differences in the positions of key ligand substituents relative to specific residues in TM3 and TM6 that are coupled to the conserved core triad of F289^{6,44}, P244^{5,50} and I155^{3,40} (Fig. 4b, d).

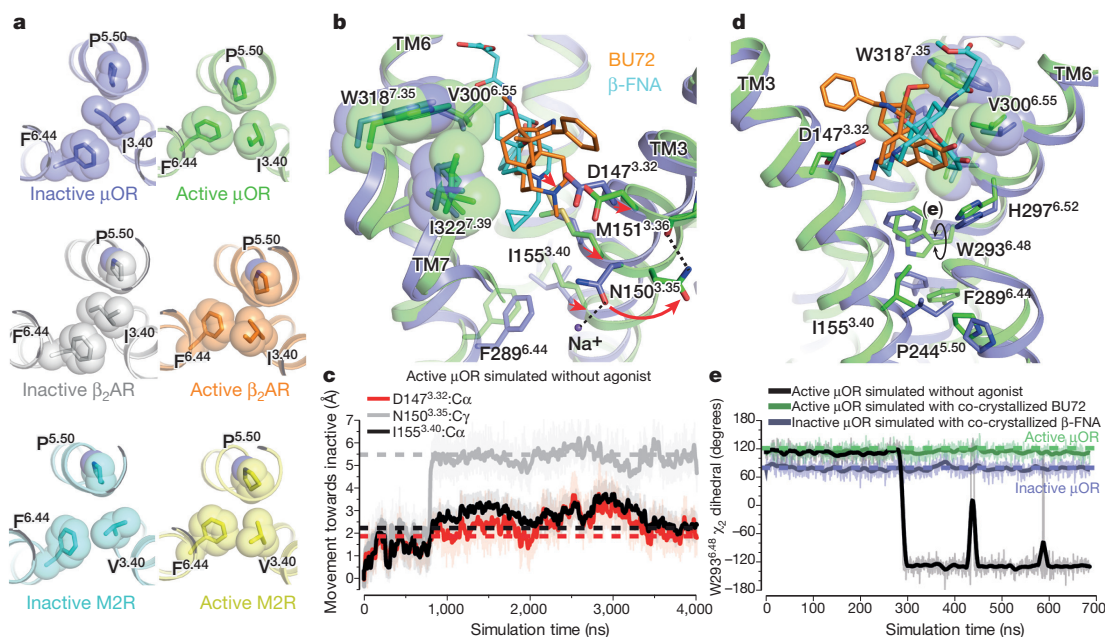


Figure 4 | Mechanisms of allosteric coupling in μ OR. **a**, Comparison of the structural rearrangements in the conserved core triad of μ OR, β_2 AR, and M2R. **b**, The morphinan ligands BU72 and β -FNA bind to the μ OR with a shared hydrophobic surface shown in spheres. BU72 binding results in a 1.5 Å displacement of TM3 towards TM2 and a rotameric change in the sodium coordinating residue N150^{3,35}. Red arrows highlight displacement of the ligand or TM3 upon activation. **c**, In a molecular dynamics simulation initiated from the active μ OR structure but with the agonist BU72 removed from the binding site, residues in TM3, including the conserved core triad residue I155^{3,40}, adopt an inactive-like conformation. The motions of I155^{3,40} and

ligand-contacting residue D147^{3,32} are tightly coupled throughout the simulation. Atom positions during simulation are plotted relative to the active structure, with positive values representing displacement towards the position in the inactive structure (see Methods). Dashed horizontal lines represent the positions of the indicated atoms in the inactive structure. **d**, W293^{6,48} is slightly closer to the phenolic aromatic of the morphinan in the active state of μ OR. **e**, Molecular dynamics simulations show that removal of the agonist BU72 from the active structure results in a change in the preferred rotamer of W293^{6,48}. Molecular dynamics results were consistent across multiple simulations; see Supplementary Information.

In the active μ OR structure, the morphinan scaffold of BU72 adopts a pose that is sterically incompatible with the inactive position of TM3. As a result, the residues of TM3 that interact with the tertiary amine (D147^{3,32}) and methyl substituent (M151^{3,36}) of the agonist shift 1.5 Å towards TM2 relative to its position in the inactive structure (Fig. 4b). We used molecular dynamics simulations to assess whether the agonist favours this displacement in TM3. In simulations of activated μ OR with the agonist removed, the previously ligand-contacting residues on TM3 quickly relax towards their inactive positions (Fig. 4c), often without global structural change of the receptor. These motions are tightly coupled to motion of the conserved core triad residue I155^{3,40} towards its inactive position, causing it to reposition relative to F289^{6,44} (Fig. 4b, c). Notably, the position of D147^{3,32} is also coupled to the rotameric state of N150^{3,35}, which coordinates an allosteric sodium ion in the inactive state and forms a hydrogen bond with a backbone carbonyl of I146^{3,31} in the active state (Fig. 4b, c).

Another link between the ligand-binding pocket and the conserved core triad may be mediated via TM6 through W293^{6,48} (Fig. 4d, e). In the active μ OR structure, the aromatic group of BU72 is positioned only slightly closer (0.6 Å) to W293^{6,48} as compared to the same aromatic group of β -FNA in the inactive state. However, molecular dynamics simulations suggest that agonists stabilize W293^{6,48} in the rotamer observed in the active-state crystal structure. In simulations of the active state, removal of BU72 results in a change in the favoured rotamer of W293^{6,48} (Fig. 4e). Conversely, we assessed whether an agonist bound to the inactive state would stabilize the side chain of W293^{6,48} in the rotamer associated with the active state. Here, we simulated the inactive-state structure with the antagonist β -FNA replaced with the agonist β -fuoxymorphamine (β -FOA), which differs from β -FNA solely in a methyl substituent at the morphinan tertiary amine³⁵ (Extended Data Fig. 7a). In these simulations, the pose of β -FOA tends to shift towards that observed for BU72 and the TM3 residues D147^{3,32} and N150^{3,35} shift towards their active-state positions; in concert, the side chain of W293^{6,48} shifts towards its active position (Extended Data Fig. 7b, c). W293^{6,48} is spatially juxtaposed to the conserved core triad residue F289^{6,44}, and the conforma-

tion of W293^{6,48} may therefore serve as an important link between the ligand-binding pocket and the triad (Fig. 4d).

Role of polar network in GPCR activation

In addition to rearrangement of the conserved core triad, comparison of the inactive and active structures of μ OR reveals an extensive network of polar interactions between the orthosteric binding pocket and the G-protein-coupling interface that must rearrange upon activation (Fig. 5a). The high-resolution electron density maps of the active μ OR allow us to detect more ordered water molecules within this polar network than have been observed in other active-state GPCR structures reported to date. To provide a better comparison of this polar network, we examined the high-resolution inactive structure of the highly homologous δ OR¹¹. The inactive structure of δ OR at a resolution of 1.8 Å identifies more water molecules within this polar network than the 2.8 Å inactive structure of μ OR and therefore highlights the contribution of many hydrogen bonds in stabilizing both the inactive and active states of opioid receptors (Fig. 5b). These hydrogen bonds represent many low-energy molecular switches that have to be broken and reformed in a concerted manner to achieve the active conformation. The polar network interactions for inactive-state μ OR are likely to be identical to the inactive state of δ OR because the specific amino acids involved in this network and their side chain conformations are identical between the two homologous receptors.

Many of the residues in the polar network are conserved (Extended Data Fig. 8a, b, e, f), suggesting that the polar network may also be conserved and play a similar role in activation of other family A GPCRs. While there are fewer water molecules observed in the active-state structure of the β_2 AR, the ones that are resolved are in the same positions as those observed in the μ OR. Even though the resolution in the active β_2 AR and M2R structures is not sufficient to observe as many ordered water molecules, the positions of the conserved side chains lining the polar core of the β_2 AR and M2R are nearly identical to those of the μ OR (Extended Data Fig. 8f), suggesting that they are stabilized by a similar hydrogen-bonding network.

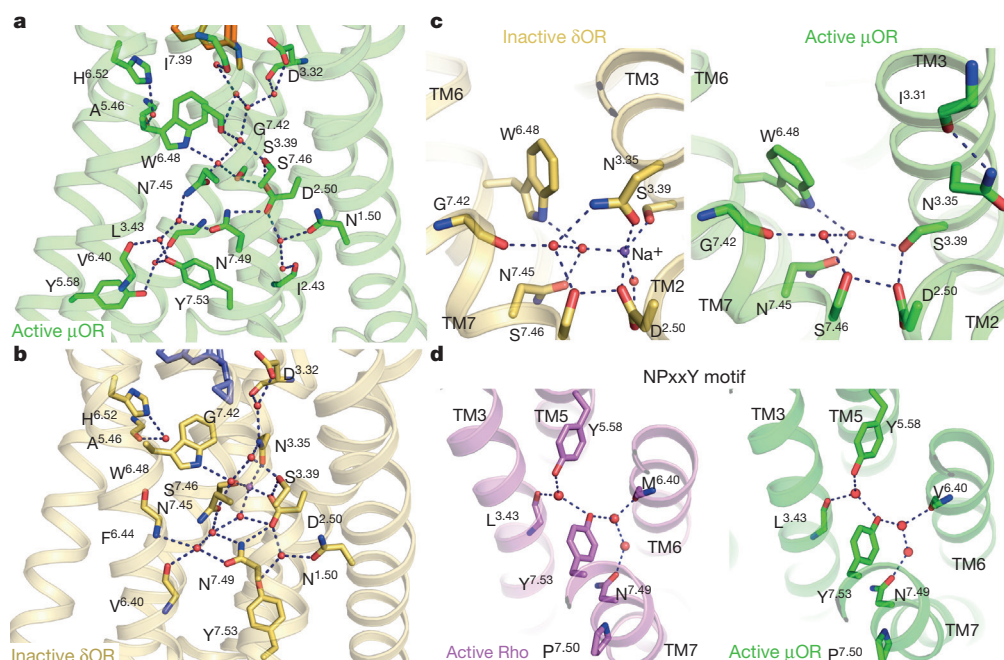


Figure 5 | Rearrangement of a conserved polar network. **a, b**, Comparison of the water-mediated polar network in the active μ OR (**a**) and high-resolution inactive δ OR (**b**) (PDB ID: 4N6H). To simplify comparisons between different receptors, only Ballesteros–Weinstein numbers are used to label amino acid side chains. The network extends from the orthosteric ligand-binding site to the

G-protein-coupling domain of the receptor. **c**, The active structure of μ OR reveals the basis for sodium ion allosteric regulation of GPCR function. Rearrangement of S3.39 and N3.35 eliminates the sodium ion coordination site in the active state. **d**, Conserved hydrogen-bonding network in the NPxxY region between active μ OR and rhodopsin.

In the high-resolution inactive-state structure of the δ OR, there is an ordered sodium ion that is adjacent to the conserved core triad and is coordinated by the side chains of D^{2.50}, N^{3.35} and S^{3.39}, as well as by W^{6.48} through a water molecule (Fig. 5b, c). A sodium ion at a similar coordination site has been observed in inactive structures of the protease-activated receptor subtype 1 (PAR1)³⁶, the adenosine A_{2A} receptor (A_{2A}R)³⁷ and the β_1 adrenergic receptor (β_1 AR)³⁸. For many GPCRs, including the μ OR, agonist-binding affinity and/or G-protein activation is allosterically inhibited by sodium³⁹. Consistent with allosteric stabilization of the inactive state by sodium, we do not observe a sodium ion in the active-state structure of the μ OR, and the residues that formed the sodium-binding site rearrange to preclude sodium binding in the active state (Fig. 5c).

The polar network ends at the cytoplasmic surface with a web of interactions involving the conserved NPxxY sequence (Fig. 5d). Upon receptor activation, the NPxxY motif in TM7 moves inward towards TM5 where N332^{7.49} and Y336^{7.53} participate in a hydrogen-bond network involving Y252^{5.58} in TM5, the backbone carbonyls of L158^{3.43} in TM3 and V285^{6.40} in TM6, and three ordered water molecules (Fig. 5d). A similar hydrogen-bond network was previously observed in the structure of metarhodopsin II¹⁹. Similar water-mediated hydrogen bonds are probably present in active β_2 AR and M2R as the side chains of key residues in this region occupy similar positions (Extended Data Fig. 3b).

It is interesting to speculate on the role of this polar network in the energetics of μ OR activation. NMR studies reveal that allosteric coupling of the agonist-binding pocket and the G-protein-binding interface in the μ OR is relatively weak, and structural changes in TM6 are observed only in the presence of a G protein mimetic nanobody (Sounier *et al.*, ref. 31). A similar observation has been made for the β_2 AR^{20–22,40}. Using both double electron–electron resonance (DEER) spectroscopy and NMR spectroscopy, we observed that in the presence of a saturating concentration of the catecholamine agonist isoproterenol only 20% of β_2 AR has TM6 in an active-like conformation²⁴. This stands in contrast to the more efficient coupling between the orthosteric binding pocket and TM6 in rhodopsin, as reflected in the ability to crystallize rhodopsin in an active state without a G protein or a G protein mimetic nanobody⁴¹, and biophysical studies that reveal a stronger allosteric coupling between the orthosteric binding pocket and TM6 in rhodopsin⁴². Comparison of the inactive-state structures of rhodopsin and the δ OR reveals that the δ OR has a more extensive polar network on the cytoplasmic side of the ligand-binding pocket (Extended Data Fig. 9). This is particularly notable when considering the network of hydrogen bonds that maintain TM6 in the inactive conformation (Extended Data Fig. 9). As noted above, a similar polar network probably stabilizes the inactive states of the μ OR and β_2 AR (Extended Data Fig. 8). The less extensive polar network stabilizing the inactive state of rhodopsin is compensated for by the covalent inverse agonist, 11-*cis*-retinal. This balance of non-covalent polar interactions and a covalent ligand imparts rhodopsin with virtually no basal activity, but the ability to rapidly and efficiently respond to photoisomerization of retinal. In contrast, the polar network in the μ OR and β_2 AR help to maintain unliganded receptors in an inactive state, but the need to disrupt this network makes activation of these receptors energetically less efficient.

Conclusion

The structure of activated μ OR presented here provides a model for how efficacy is encoded within small chemical differences in otherwise structurally similar morphinan ligands. Comparison of active μ OR with other active-state structures offers insights into shared and subtype-specific mechanisms for the activation of family A GPCRs. Additionally, the extensive reorganization of the polar network required to achieve the fully active state may explain the inefficient allosteric coupling of the orthosteric pocket and the G-protein-coupling interface observed in NMR studies for both the μ OR and

the β_2 AR. It is possible that subtle ligand-specific differences in the polar network connections may contribute to preferential activation of different signalling proteins.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 March; accepted 30 June 2015.

Published online 5 August 2015.

- Matthes, H. W. *et al.* Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ -opioid-receptor gene. *Nature* **383**, 819–823 (1996).
- Brownstein, M. J. A brief history of opiates, opioid peptides, and opioid receptors. *Proc. Natl Acad. Sci. USA* **90**, 5391–5393 (1993).
- Schumacher, M. A., Basbaum, A. I. & Naidu, R. K., (McGraw-Hill Medical, 2015).
- Raehal, K. M., Walker, J. K. & Bohn, L. M. Morphine side effects in β -arrestin 2 knockout mice. *J. Pharmacol. Exp. Ther.* **314**, 1195–1201 (2005).
- Bohn, L. M., Gainetdinov, R. R., Lin, F.-T., Lefkowitz, R. J. & Caron, M. G. μ -Opioid receptor desensitization by β -arrestin-2 determines morphine tolerance but not dependence. *Nature* **408**, 720–723 (2000).
- Bohn, L. M. *et al.* Enhanced morphine analgesia in mice lacking β -arrestin 2. *Science* **286**, 2495–2498 (1999).
- Pasternak, G. W. & Pan, Y.-X. Mu opioids and their receptors: evolution of a concept. *Pharmacol. Rev.* **65**, 1257–1317 (2013).
- Chavkin, C. & Goldstein, A. Specific receptor for the opioid peptide dynorphin: structure–activity relationships. *Proc. Natl Acad. Sci. USA* **78**, 6543–6547 (1981).
- Manglik, A. *et al.* Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
- Granier, S. *et al.* Structure of the δ -opioid receptor bound to naltrindole. *Nature* **485**, 400–404 (2012).
- Fenalti, G. *et al.* Molecular control of δ -opioid receptor signalling. *Nature* **506**, 191–196 (2014).
- Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180 (2011).
- Ring, A. M. *et al.* Adrenaline-activated structure of β_2 -adrenoceptor stabilized by an engineered nanobody. *Nature* **502**, 575–579 (2013).
- Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2 -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β_2 -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Haga, K. *et al.* Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482**, 547–551 (2012).
- Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Choe, H.-W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- β_2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).
- Nygaard, R. *et al.* The dynamic process of β_2 -adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
- Manglik, A. & Koblika, B. The role of protein dynamics in GPCR function: insights from the β_2 AR and rhodopsin. *Curr. Opin. Cell Biol.* **27**, 136–143 (2014).
- Manglik, A. *et al.* Structural insights into the dynamic process of β_2 -adrenergic receptor signalling. *Cell* **161**, 1101–1111 (2015).
- De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled β -adrenergic receptor. *J. Biol. Chem.* **255**, 7108–7117 (1980).
- Rasmussen, S. G. F. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
- Schiller, P. W. *et al.* Synthesis and *in vitro* opioid activity profiles of DALDA analogues. *Eur. J. Med. Chem.* **35**, 895–901 (2000).
- Neilan, C. L. *et al.* Characterization of the complex morphinan derivative BU72 as a high efficacy, long-lasting μ -opioid receptor agonist. *Eur. J. Pharmacol.* **499**, 107–116 (2004).
- Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* **38**, 29–51 (2009).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Wu, H. *et al.* Structure of the human κ -opioid receptor in complex with JDTic. *Nature* **485**, 327–332 (2012).
- Sounier, R. *et al.* Propagation of conformational changes during μ -opioid receptor activation. *Nature* <http://dx.doi.org/10.1038/nature14680> (2015).
- Chaturvedi, K., Shahrestanifar, M. & Howells, R. D. μ Opioid receptor: role for the amino terminus as a determinant of ligand binding affinity. *Brain Res. Mol. Brain Res.* **76**, 64–72 (2000).
- Gales, C. *et al.* Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. *Nature Struct. Mol. Biol.* **13**, 778–786 (2006).

34. Husbands, S. M. *et al.* BU74, a complex oripavine derivative with potent kappa opioid receptor agonism and delayed opioid antagonism. *Eur. J. Pharmacol.* **509**, 117–125 (2005).
35. Takemori, A. E., Larson, D. L. & Portoghese, P. S. The irreversible narcotic antagonistic and reversible agonistic properties of the fumaramate methyl ester derivative of naltrexone. *Eur. J. Pharmacol.* **70**, 445–451 (1981).
36. Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387–392 (2012).
37. Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
38. Miller-Gallacher, J. L. *et al.* The 2.1 Å resolution structure of cyanopindolol-bound β_1 -adrenoceptor identifies an intramembrane Na^+ ion that stabilises the ligand-free receptor. *PLoS ONE* **9**, e92727 (2014).
39. Pert, C. B., Pasternak, G. & Snyder, S. H. Opiate agonists and antagonists discriminated by receptor binding in brain. *Science* **182**, 1359–1361 (1973).
40. Manglik, A. *et al.* Structural insights into the dynamic process of β_2 -adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
41. Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H. W. & Ernst, O. P. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454**, 183–187 (2008).
42. Knierim, B., Hofmann, K. P., Gartner, W., Hubbell, W. L. & Ernst, O. P. Rhodopsin and 9-demethyl-retinal analog: effect of a partial agonist on displacement of transmembrane helix 6 in class A G protein-coupled receptors. *J. Biol. Chem.* **283**, 4967–4974 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge support from the Stanford Medical Scientist Training Program and the American Heart Association (A.M.), National Institutes of Health grants R37DA036246 (B.K.K. and S.G.) and R01GM083118 (B.K.K.), a Terman Faculty Fellowship (R.O.D.), Eli Lilly and Company through the Lilly Research Program (R.O.D.), and the Mathers Foundation (B.K.K. and W.I.W.). We also acknowledge the

National Institute of Drug Abuse Drug Supply Program for providing [Dmt¹]DALDA. We thank D. Maurel and S. Agnel from the ARPEGE facility (Institut de Génomique Fonctionnelle) for assistance with cell-based G_i coupling assays, H. El Hassan for expert technical assistance, and S. Hertig, N. Latorraca and K. Cavalotti for assistance with molecular dynamics simulations and analysis.

Author Contributions W.H. developed functional purification protocols, expressed and purified μOR , characterized the effect of nanobodies and G_i on μOR ligand affinity, identified Nb39 for crystallography of the μOR -Nb complex, performed crystallization trials, data collection, structure determination and refinement. A.M. established the project with biochemistry of active μOR , prepared samples for llama immunization, validated nanobody activity, performed crystallization trials, and identified initial crystals of the μOR -BU72-Nb complex suitable for diffraction studies. A.J.V. analysed the polar network. A.J.V., E.F. and A.S. performed and analysed molecular dynamics simulations with supervision from R.O.D. T.L. identified μOR -binding nanobodies with supervision from J.S. S.G. established the biochemistry for purification of agonist-bound μOR and prepared samples for μOR immunization. H.E.K. helped with data collection and processing. T.S.T. helped with the characterization of the amino-terminal region. R.K. and P.G. analysed BU72 and assessed alternative ligand structures. S.M.H. synthesized BU72. K.E.L. and J.R.T. helped with selection of opioid ligands including BU72 and performed dissociation kinetics experiments. W.I.W. supervised structure refinement. A.M. and B.K.K. provided overall project supervision and wrote the manuscript with W.H. and R.O.D.

Author Information Coordinates and structure factors for the μOR -BU72-Nb39 complex have been deposited in the Protein Data Bank under accession code 5C1M. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (amanglik@stanford.edu) or B.K.K. (kobilka@stanford.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Expression and purification of μ OR. Full length *Mus musculus* μ OR bearing an amino-terminal Flag epitope tag and a carboxy-terminal $6 \times$ His tag was expressed in *Sf9* insect cells using the BestBac baculovirus system (Expression Systems). To facilitate removal of flexible amino- and carboxy-terminal regions, a tobacco etch virus protease recognition sequence was inserted after residue 51 and a rhinovirus 3C protease recognition sequence was inserted before residue 359. Insect cells were infected with baculovirus encoding μ OR at a density of 4×10^6 cells ml^{-1} for 48–60 h at 27 °C. Receptor was solubilized and purified in a final buffer comprised of 25 mM HEPES pH 7.4, 100 mM NaCl, 0.01% MNG (Anatrace), and 0.001% cholesterol hemisuccinate (CHS), as previously described⁹.

Llama immunization and selection of μ OR-binding nanobodies. Purified μ OR bound to the antagonist naloxone and with the amino and carboxy termini cleaved was incubated with an excess of the agonist [Dmt¹]DALDA (NIDA Drug Supply Program), and further purified by size-exclusion chromatography in a buffer comprised of 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1% dodecylmaltoside (DDM, Anatrace), 0.03% CHAPS, 0.01% CHS, and 1 μ M [Dmt¹]DALDA. The resulting agonist-bound receptor was reconstituted into phospholipid vesicles composed of DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine, Avanti Polar Lipids) and Lipid A in a 10:1 (w:w) ratio at a final receptor concentration of 1.3 mg ml^{-1} . The resulting reconstituted receptor was flash frozen in liquid nitrogen in 100 μ g aliquots for llama immunization.

One llama was immunized over a period of 6 weeks with 0.3 mg of liposome-reconstituted μ OR purified as described above and bound to the agonist [Dmt¹]DALDA. A phage display library of nanobodies was prepared from peripheral blood lymphocytes as previously described⁴³. μ OR-binding nanobodies were identified by selecting phages that bound liposome-reconstituted μ OR in the presence or absence of agonist. Ten clones from three families were enriched during rounds of selection. One family, which includes Nb39, bind the intracellular surface and function as G-protein mimetics. Another family of nanobodies, including clone Nb35, was identified that competes directly with orthosteric antagonists of the μ OR and binds at the extracellular surface.

Expression and purification of nanobodies. Nanobodies bearing a carboxy-terminal $6 \times$ His tag were expressed in the periplasm of *Escherichia coli* strain WK6. Cultures were grown to an OD₆₀₀ of 1.0 at 37 °C in Terrific Broth medium containing 0.1% glucose, 2 mM MgCl₂, and 50 μ g ml^{-1} ampicillin and induced with 0.5 mM isopropyl- β -D-thiogalactoside (IPTG). Cells were harvested after overnight growth at 25 °C and incubated in a buffer containing 200 mM Tris pH 8.0, 0.5 mM EDTA, 500 mM sucrose and 0.5 mg ml^{-1} lysozyme for 1 h at 25 °C. Bacteria were osmotically lysed by rapid dilution in a 4 \times volume of water. The periplasmic fraction was isolated by centrifugation of cell debris, and was supplemented with NaCl to a final concentration of 150 mM as well as imidazole to a final concentration of 25 mM. Nanobodies were purified from the periplasmic fraction by nickel affinity chromatography, and were subsequently purified by size-exclusion chromatography in a buffer comprised of 25 mM HEPES pH 7.5 and 100 mM NaCl. Peak fractions were pooled and concentrated to approximately 5 mM.

Radioligand characterization of BU72 and Nb39. Radioligand competition assays were performed using purified μ OR reconstituted into high-density lipoprotein (HDL) particles comprised of the lipids POPC and POPG (Avanti Polar Lipids) in a 3:2 molar ratio as previously described^{44,45}. Heterotrimeric G_i was prepared by co-expressing human G α 1, human G β 1 and G γ 2 subunits in Hi5 insect cells using baculoviruses encoding the individual subunits. The G protein was purified as previously described²⁵.

For competition binding experiments, a mixture of 0.02 nM μ OR and 2.3 nM ³H-diprenorphine (³H-DPN, Perkin Elmer) was incubated with varying concentrations of agonist in a binding buffer comprised of 25 mM HEPES pH 7.4, 100 mM NaCl, and 0.1% BSA. Experiments were also performed with either 5 μ M Nb39 or 500 nM G_i. Binding reactions were incubated for 4 h at 25 °C. Free radioligand was separated from bound radioligand by rapid filtration onto a Whatman GF/B filter pretreated with 0.1% polyethylenimine with the aid of a 48-well harvester (Brandel). Radioligand activity was measured by liquid scintillation counting. Competition binding data were fit to a one-site model for μ OR alone and μ OR with Nb39 and a two-site model for μ OR incubated with G_i using GraphPad Prism 6.0.

Dissociation studies for BU72 were performed using the method of Motulsky and Mahan⁴⁶. ³H-DPN was diluted in an assay buffer comprised of 20 mM Tris, pH 7.4, 150 mM NaCl, and 0.05% BSA containing μ OR in HDL particles with

either vehicle or different concentrations of BU72 alone or in the presence of Nb39. Binding reactions were incubated at 25 °C in the dark and nonspecific binding was determined in the presence of 10 μ M naloxone. Aliquots of this binding reaction were removed at specified time points over the course of 2 or 8 h and filtered through Whatman GF/C filters with the aid of a Brandel harvester. As above, radioligand activity was measured by liquid scintillation counting. Dissociation rates for BU72 were determined by fitting data in the 'kinetics of competitive binding' program in GraphPad Prism 6.02. For K1 and K2, rates of ³H-DPN association and dissociation were determined through independent studies following the same method as above.

Purification and crystallization of the μ OR–BU72–Nb39 complex. Initial crystals of a μ OR–BU72–Nb complex diffracted to low resolution. In an effort to improve the specific activity of purified μ OR, we used an extracellular binding nanobody, Nb35 during the purification of the receptor. Nb35 binds in the orthosteric pocket of μ OR with a K_i of 12 nM. Purified ligand-free μ OR, with the amino and carboxy termini cleaved was mixed with a 2.5 \times excess of Nb35, and the complex was isolated by Ni-NTA chromatography. Functional μ OR in complex with Nb35 was eluted in a buffer containing 25 mM HEPES pH 7.5, 100 mM NaCl, 0.01% MNG, 0.001% CHS, and 250 mM imidazole. The agonist BU72 and Nb39 were then added in excess. In the presence of Nb39, BU72 has exceptionally high affinity for the μ OR and displaces Nb35, resulting in the desired μ OR–BU72–Nb39 complex. The μ OR–BU72–Nb39 complex was isolated by size-exclusion chromatography and concentrated to approximately 50 mg ml^{-1} for crystallization trials.

For crystallization, the purified μ OR–BU72–Nb39 complex was reconstituted into a 10:1 (w:w) mixture of monoolein and cholesterol (Sigma). The protein solution and lipid were mixed in a 1:1.5 ratio (w:w), and the lipidic cubic phase was attained using the two-syringe method⁴⁷. Thirty nanolitres of the resulting mesophase was dispensed onto 96-well glass plates and overlaid with 500 nl of precipitant solution using a Gryphon LCP robot (Art Robbins Instruments). Crystals grew in a precipitant comprised of 15–25% PEG300, 100 mM HEPES pH 7.0–7.5, 1% 1,2,3-heptanetriol, 0.5–1.0% polypropylene glycol P 400 (Hampton Research) and 100–300 mM (NH₄)₂HPO₄. Crystals were observed after 2 days and reached full size in 1 week. Crystals were harvested with mesh grid loops (MiTeGen) and flash frozen in liquid nitrogen. Loops were screened at the Advanced Photon Source GM/CA beamlines 23ID-B and 23ID-D. Diffraction data were collected at a wavelength of 1.033 Å using a 10 μ m beam with fivefold attenuation and exposed for 0.5–1 s. An oscillation width of 0.1–0.5° was used and diffraction images from four crystals were merged to create the final data set.

Structure determination and refinement. Diffraction images were indexed, scaled, and merged by XDS⁴⁸ with statistics summarized in Extended Data Table 1. The structure was determined by molecular replacement in Phaser⁴⁹ with initial search models including inactive μ OR (PDB ID: 4DKL) and Nb80 (PDB ID: 3P0G). The structural model was iteratively rebuilt in Coot⁵⁰ and refined using Phenix⁵¹. Final refinement statistics are summarized in Extended Data Table 1. Molprobit⁵² was used for structure analysis and indicated that 96.1% of residues were within favoured Ramachandran regions with 3.9% in allowed regions. No residues were identified as Ramachandran outliers. Figures were prepared using MacPyMOL (Schrödinger, Inc.).

System setup for molecular dynamics simulations. Simulations of the μ OR were based on both the antagonist-bound inactive-state crystal structure (PDB ID: 4DKL) and the agonist-bound active-state crystal structure described in this manuscript. Coordinates were prepared by first removing all non-ligand and non-receptor molecules except for the cholesterol neighbouring TM7 and for crystallographic water molecules near the receptor. For inactive μ OR simulations, the T4 lysozyme was removed and acetyl and methylamide capping groups were placed on R263^{ICL3} and E270^{ICL3}. For active μ OR simulations, the nanobody was removed. In both cases, Prime (Schrödinger, Inc.) was used to model missing side-chains, and capping groups were then added to the N and C termini of the receptor. Histidine residues were simulated as the neutral H₂ tautomer. Other titratable residues were simulated in their dominant protonation state at pH 7 except for D114^{2,50}, which was charged in inactive simulations and neutral in active simulations. A sodium ion was placed adjacent to D114^{2,50} in inactive simulations.

The μ OR was simulated in seven distinct conditions. These include: (1) the unliganded, inactive μ OR, prepared by deleting the covalently bound, co-crystallized ligand, β -FNA, and adding a proton in its place to K233^{5,39}; (2) the inactive μ OR with the co-crystallized ligand β -FNA; (3) the inactive μ OR with agonist β -FOA (which does not bind covalently); (4) the unliganded, active μ OR, prepared by deleting the co-crystallized ligand, BU72; (5) the active μ OR with the co-crystallized ligand BU72; (6) the active μ OR with the co-crystallized ligand BU72, with the N-terminal peptide deleted; and (7) the active μ OR with the

antagonist BU74, with the N-terminal peptide deleted. Simulations of the active μ OR without an N-terminal peptide were prepared by deleting residues 52 through 64 of the receptor. Simulations with β -FOA were prepared by docking β -FOA to the crystallographic pose of β -FNA. Simulations with BU74 were prepared by docking BU74 to the crystallographic pose of BU72 and rotating the torsion angle of the methylcyclopropyl group to agree with that of β -FNA's methylcyclopropyl group.

We performed three to six simulations per condition (Supplementary Information). Simulations in a given condition were initiated from identical structures, but with initial atom velocities assigned independently and randomly.

It should be noted that in all liganded simulations, including those with β -FNA, β -FOA, BU72, and BU74, the ligand's tertiary amine nitrogen was protonated and therefore the ligand was simulated as a cation. This is necessary for the ligand to form the conserved salt bridge with neighbouring D147^{3,32}.

Each of the resulting prepared μ OR structures was then aligned to the orientations of proteins in membranes (OPM)⁵³ entry for the inactive μ OR using MacPyMOL (Schrödinger, Inc.). The μ OR was modified with disulphide bridges and inserted into a hydrated, equilibrated palmitoylcholinephosphatidylcholine (POPC) bilayer using the CHARMM-GUI interface^{54–57}. Sodium and chloride ions were added to neutralize the system, reaching a final concentration of approximately 150 mM. All simulations contained one μ OR embedded in a lipid bilayer with 160 POPC molecules.

Simulation protocol and analysis. Each simulation was performed on two GPUs using the CUDA version of PMEMD (Particle Mesh Ewald Molecular Dynamics) in Amber14^{58–60} with 2.5-fs time steps. Simulations were heated from 0 K to 100 K in the NVT ensemble and then from 100 K to 310 K in the NPT ensemble, both with 10.0 kcal mol^{−1} Å^{−2} harmonic restraints applied to the protein and to the lipids. Initial velocities were assigned randomly at the first heating step with Langevin dynamics. Subsequently, simulations were equilibrated in the NPT ensemble at 310 K (controlled with a Langevin thermostat) with pressure of 1 bar (controlled with anisotropic Berendsen weak-coupling barostat), with harmonic restraints on all protein atoms tapered off 1.0 kcal mol^{−1} Å^{−2} in a stepwise fashion every 2.0 ns starting at 5.0 kcal mol^{−1} Å^{−2} to 0.0 kcal mol^{−1} Å^{−2}, for a total of 12.0 ns of equilibration. Bond lengths to hydrogen atoms were constrained using SHAKE. Production simulations were performed in the NPT ensemble at 310 K and 1 bar, using a Langevin thermostat for temperature coupling and a Monte Carlo barostat for pressure coupling, and were initiated from the final snapshot of the corresponding equilibration simulation. Non-bonded interactions were cut off at 9.0 Å, and long-range electrostatic interactions were computed using the particle mesh Ewald (PME) method, with an Ewald coefficient β of approximately 0.31 Å^{−1} and B-spline interpolation of order 4. The FFT grid size was 84 × 84 × 96 for all simulations. We performed a total of 27 simulations, which are summarized in the Supplementary Information.

We used the CHARMM36 parameter set for protein molecules, lipid molecules, and salt ions, and the CHARMM TIP3P model for water; protein parameters incorporated CMAP terms^{61–64}. These parameters were assigned using the ParmEd implementation of Chamber, provided with AmberTools14. Ligand parameters were based on the results from the CHARMM ParamChem web server, version 0.9.7.1⁶⁵, and incorporated specific modifications to partial charges. When ParamChem reported large penalties (errors) for estimated partial charges, better estimates were determined by submitting appropriately chosen fragments to ParamChem (see below).

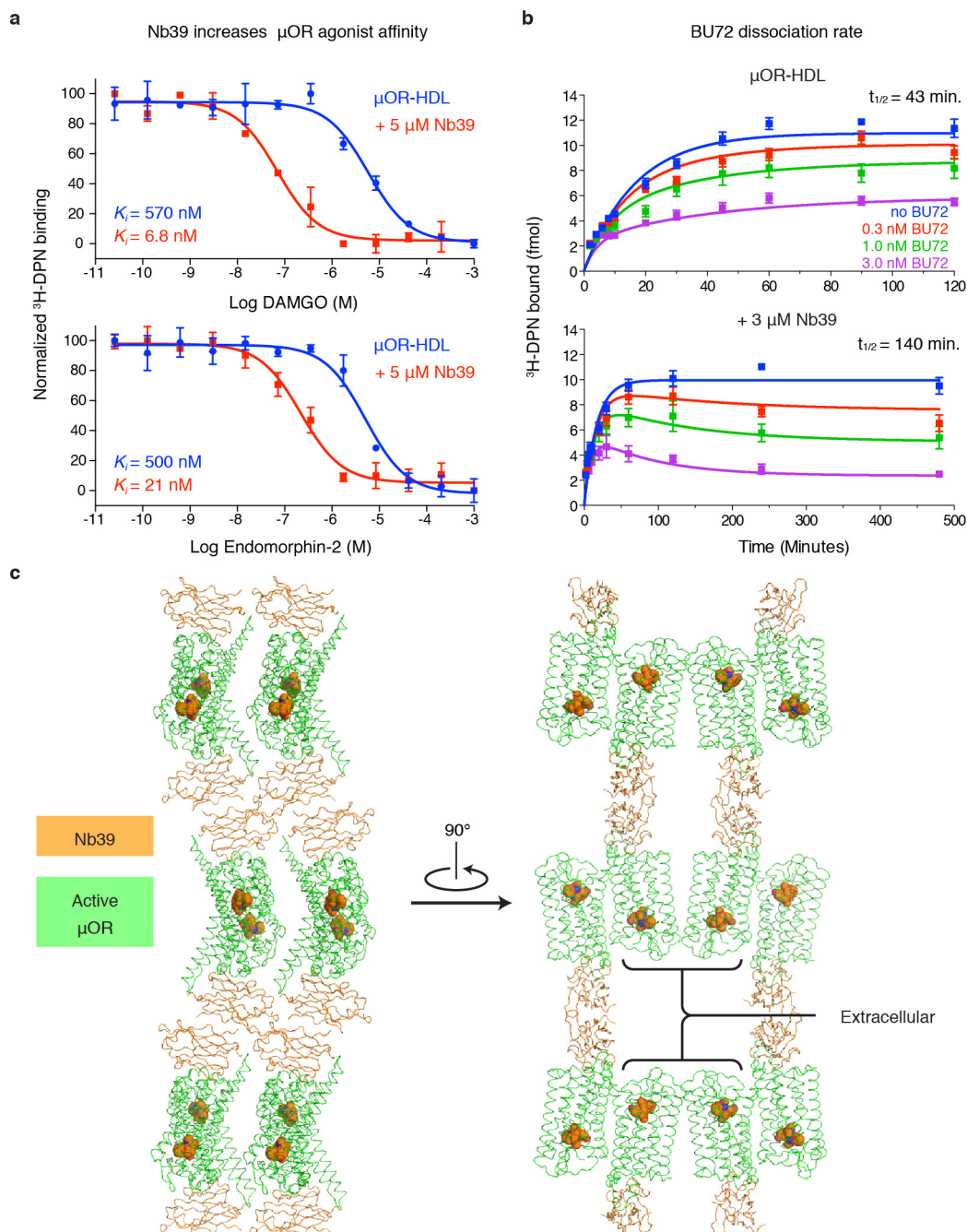
Parameterization of β -FNA, the co-crystallized ligand in the inactive μ OR structure, required additional steps due to the added complexity of its covalent bond with K233^{5,39} on the receptor. To assign parameters to the rest of β -FNA, the molecule was fragmented as shown in Supplementary Information section “Parameterization” and then re-assembled. In addition, a molecule consisting of K233^{5,39} bonded to the portion of β -FNA on the extracellular side of its morphinan scaffold was submitted to ParamChem, and the results were used to modify charges near the bond between the β -FNA carbon and Lys side-chain nitrogen. A modified version of the ParmEd program was used to form a bond between β -FNA and K233^{5,39}, to parameterize that and neighbouring bonds, and to modify neighbouring partial charges according to the derived parameters. The related ligand β -FOA (which does not bind covalently to the μ OR) was divided into the same three fragments as β -FNA for parameterization, with the following modifications: the bond between the ligand and K233^{5,39} was cleaved; a C=C double bond was formed between the two carbons adjacent to this bond; and the methylcyclopropyl group was replaced with a methyl group. The co-crystallized ligand

of the active μ OR, BU72, was similarly parameterized by splitting it into fragments, individually uploading the fragments to the ParamChem web server, and assembling the results together. The fragments can be found in the Supplementary Information. The ligand, BU74, was parameterized identically to BU72, except that its tertiary amine methyl group was replaced with a methylcyclopropyl group.

Quantitative analysis of trajectories was conducted in both cpptraj, an Amber14 software package, and VMD⁶⁶. Plots were smoothed with a moving average, with a symmetric triangular smoothing window. The window length was 20 ns, except during the first 10 ns of each simulation, when the smoothing window extended from time zero to double the time at which the data was recorded. Plots were rendered in R with the ggplot2 package as well as in Python with the Matplotlib package.

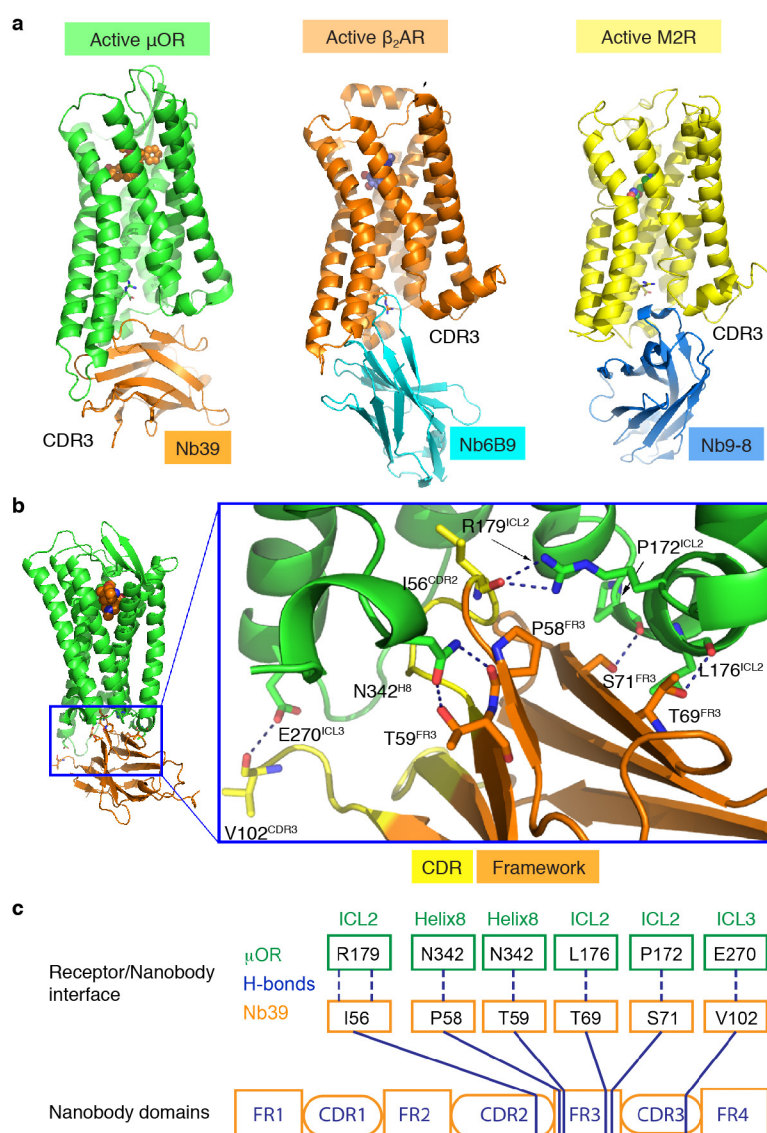
To compute movement of specified atoms from the active position towards the inactive position, as plotted in Fig. 4c, the active and inactive crystal structures and each simulation frame were aligned using the C α atoms of residues 80–95, 102–13, and 181–205 (TMs 1, 2 and 4). For each specified atom, the atom position in simulation was projected onto the line connecting the atom positions in the active and inactive crystal structures, and the distance of the projected point from the position in the active crystal structure position was determined.

43. Pardon, E. *et al.* A general protocol for the generation of nanobodies for structural biology. *Nature Protocols* **9**, 674–693 (2014).
44. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
45. Kuszak, A. J. *et al.* Purification and functional reconstitution of monomeric mu-opioid receptors: allosteric modulation of agonist binding by Gi2. *J. Biol. Chem.* **284**, 26732–26741 (2009).
46. Motulsky, H. J. & Mahan, L. C. The kinetics of competitive radioligand binding predicted by the law of mass action. *Mol. Pharmacol.* **25**, 1–9 (1984).
47. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
48. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
49. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
50. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
51. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
52. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
53. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).
54. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
55. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE* **2**, e880 (2007).
56. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
57. Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).
58. Case, D. A. *et al.* AMBER 14. (University of California, San Francisco, 2014).
59. Le Grand, S., Götz, A. W. & Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).
60. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with Amber on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
61. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
62. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
63. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
64. MacKerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
65. Vanommeslaeghe, K. *et al.* CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
66. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).



Extended Data Figure 1 | Characterization of Nb39 and lattice interactions in μOR -Nb39 crystals. **a**, ^3H -diprenorphine (^3H -DPN) competition binding shows increased affinity for μOR -selective agonists DAMGO and endomorphin-2 in the presence of Nb39. **b**, The dissociation half-life ($t_{1/2}$) of BU72 was determined by measuring the association rate of the antagonist ^3H -DPN in the presence of the indicated concentrations of BU72. The

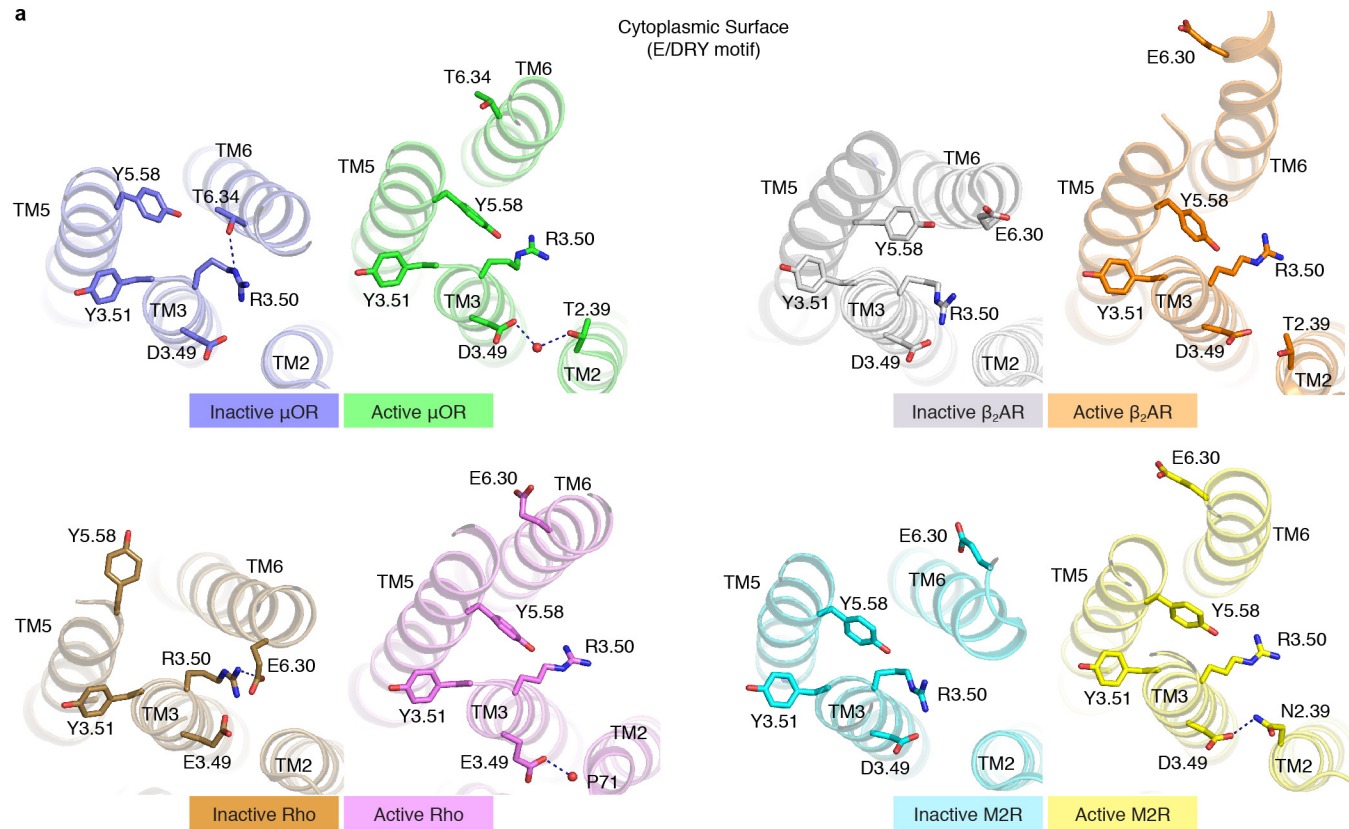
dissociation $t_{1/2}$ of BU72 is 43 min and increases to 140 min in presence of Nb39. Panels **a** and **b** are representative of at least three experiments performed in triplicate, and the data and error bars represent the mean \pm s.e.m. **c**, Crystal lattice packing of the μOR -Nb39 complex shows that most of the contacts are mediated by Nb39. The μOR extracellular domain is not involved in any contacts.



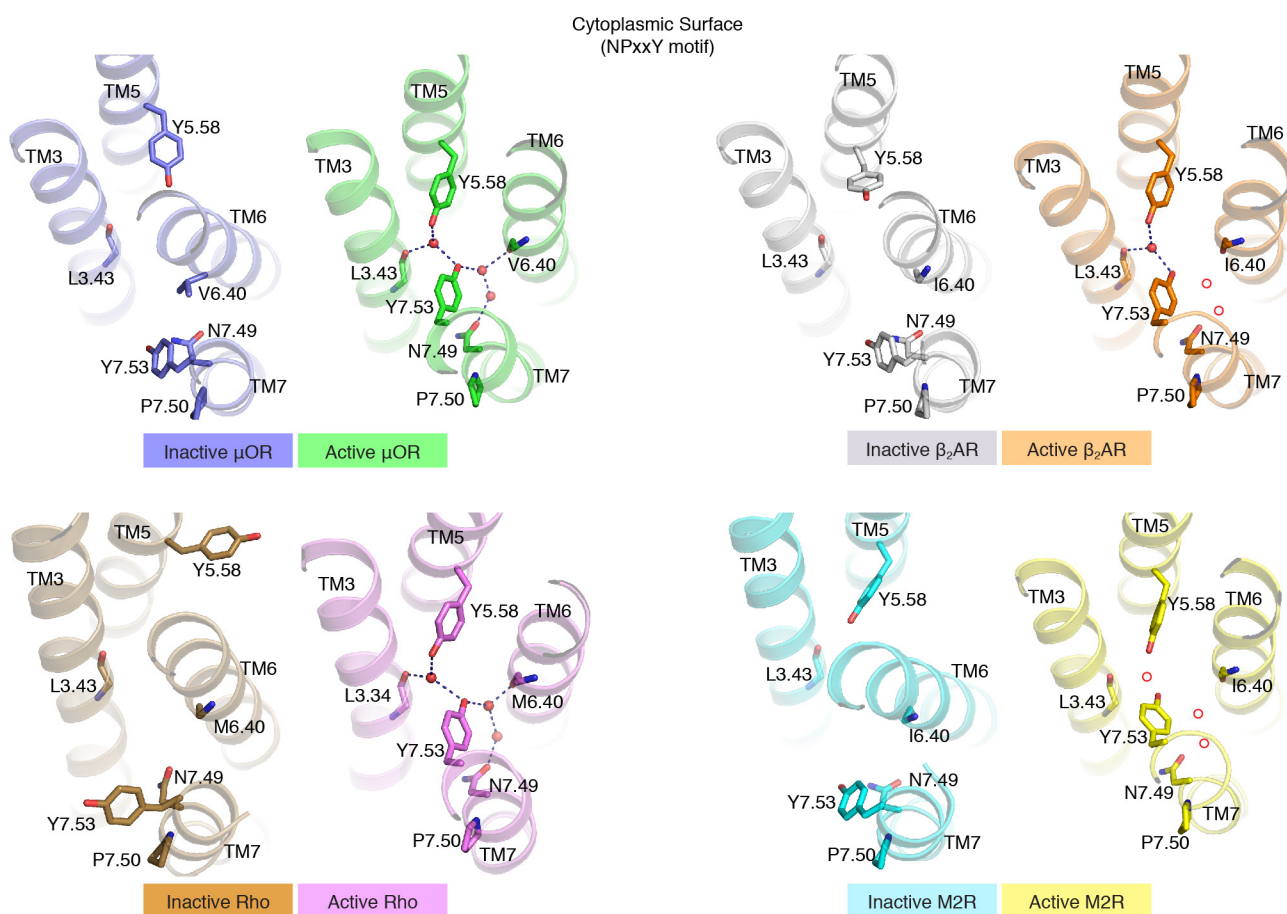
Extended Data Figure 2 | μ OR–Nb39 interface. **a**, Nb39 does not penetrate as deeply into the core of the μ OR when compared with the β_2 AR–Nb80 complex and the M2R–Nb9-8 complex. In the β_2 AR–Nb80 and M2R–Nb9-8 complexes, nanobody CDR3 residues bind within the core of the receptor transmembrane bundle. In comparison, Nb39 binding involves more

framework residues. Notably, seven residues of CDR3 remained unresolved in the final model of the μ OR–Nb39 complex. **b**, Nb39 interacts primarily through hydrogen bonds with residues from ICL2, ICL3 and helix 8 of the μ OR. **c**, Schematic representation of the interactions between μ OR and Nb39 highlighting the numerous Nb39 framework interactions.

a



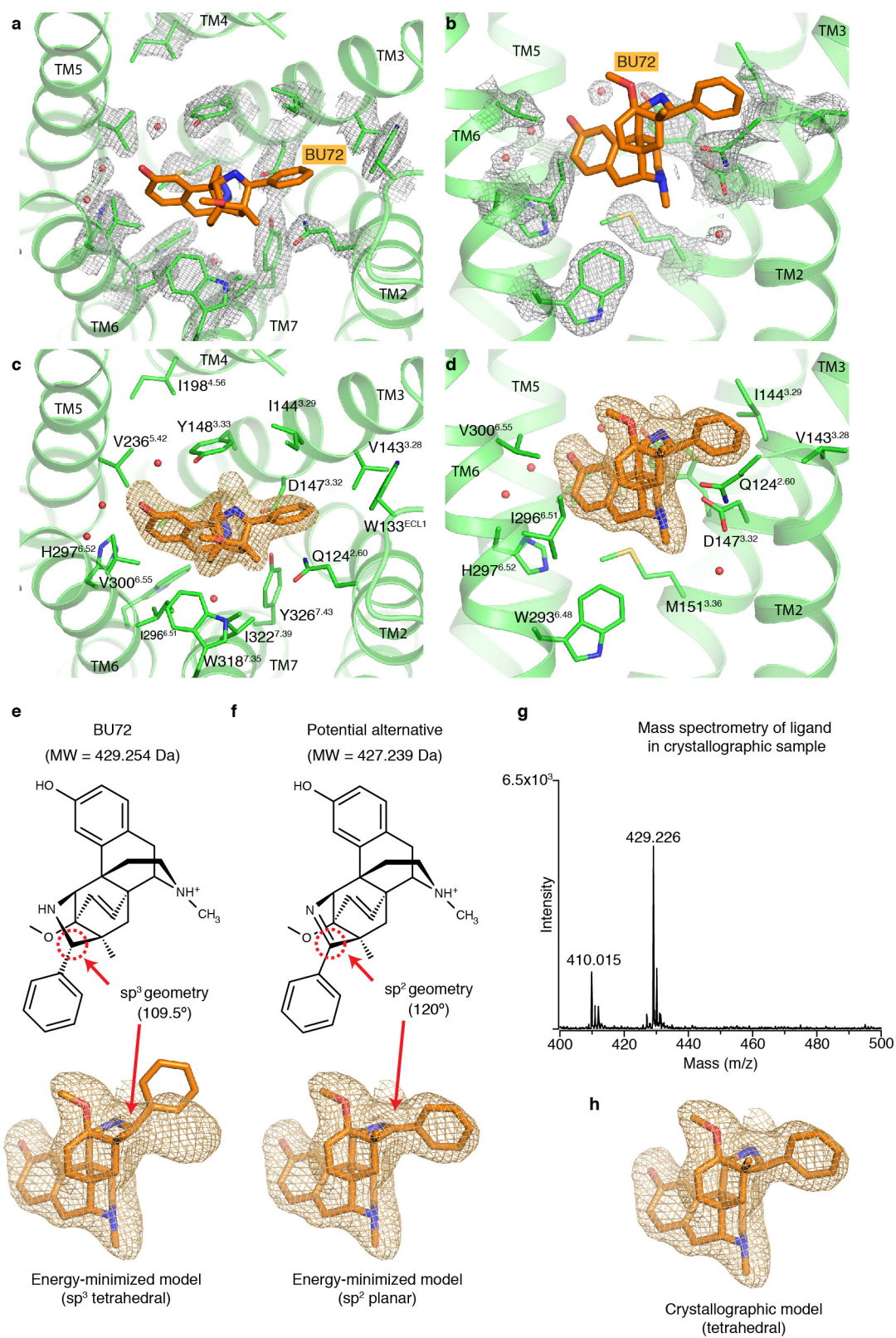
b



Extended Data Figure 3 | Cytoplasmic domain rearrangements in conserved regions.

a. The E/DRY sequence is a highly conserved motif within family A GPCRs responsible for constraining receptors in an inactive conformation. Comparisons of inactive- and active-state structures around the conserved E/DRY residues at the cytoplasmic surface of the μ OR, the M2 muscarinic receptor (M2R), the β_2 adrenergic receptor (β_2 AR) and rhodopsin (Rho) are shown here. Hydrogen bonds are shown as dotted lines. **b.** The

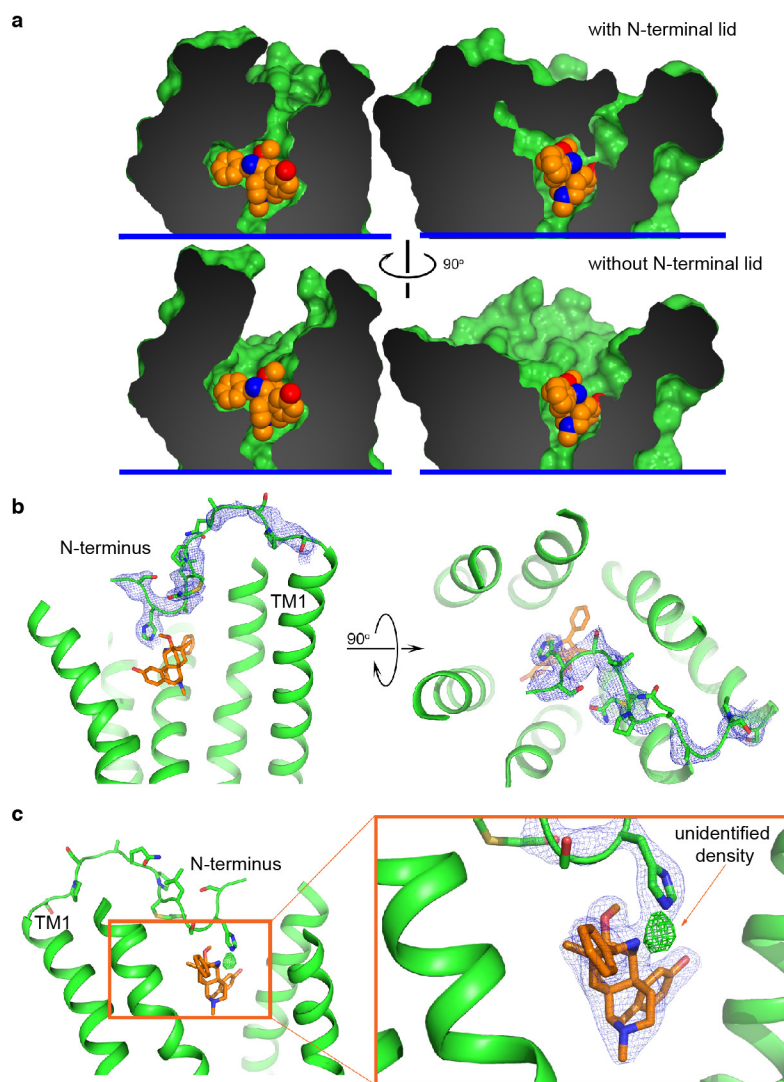
NPxxY motif is a highly conserved sequence in TM7 among family A GPCRs. In the active state μ OR, Y^{7.53} and N^{7.49} in TM7 interact with Y^{5.58} in TM5 and the backbone carbonyl of L^{3.43} in TM3 through a water-mediated polar network. A similar network is observed in the active state of rhodopsin. While waters are not observed in the lower-resolution structures of the β_2 AR and M2R, the positions of the side chains of Y^{7.53}, N^{7.49} and Y^{5.58} suggest a similar water-mediated network with putative waters represented by red circles.



Extended Data Figure 4 | Conformation of the binding pocket and BU72.

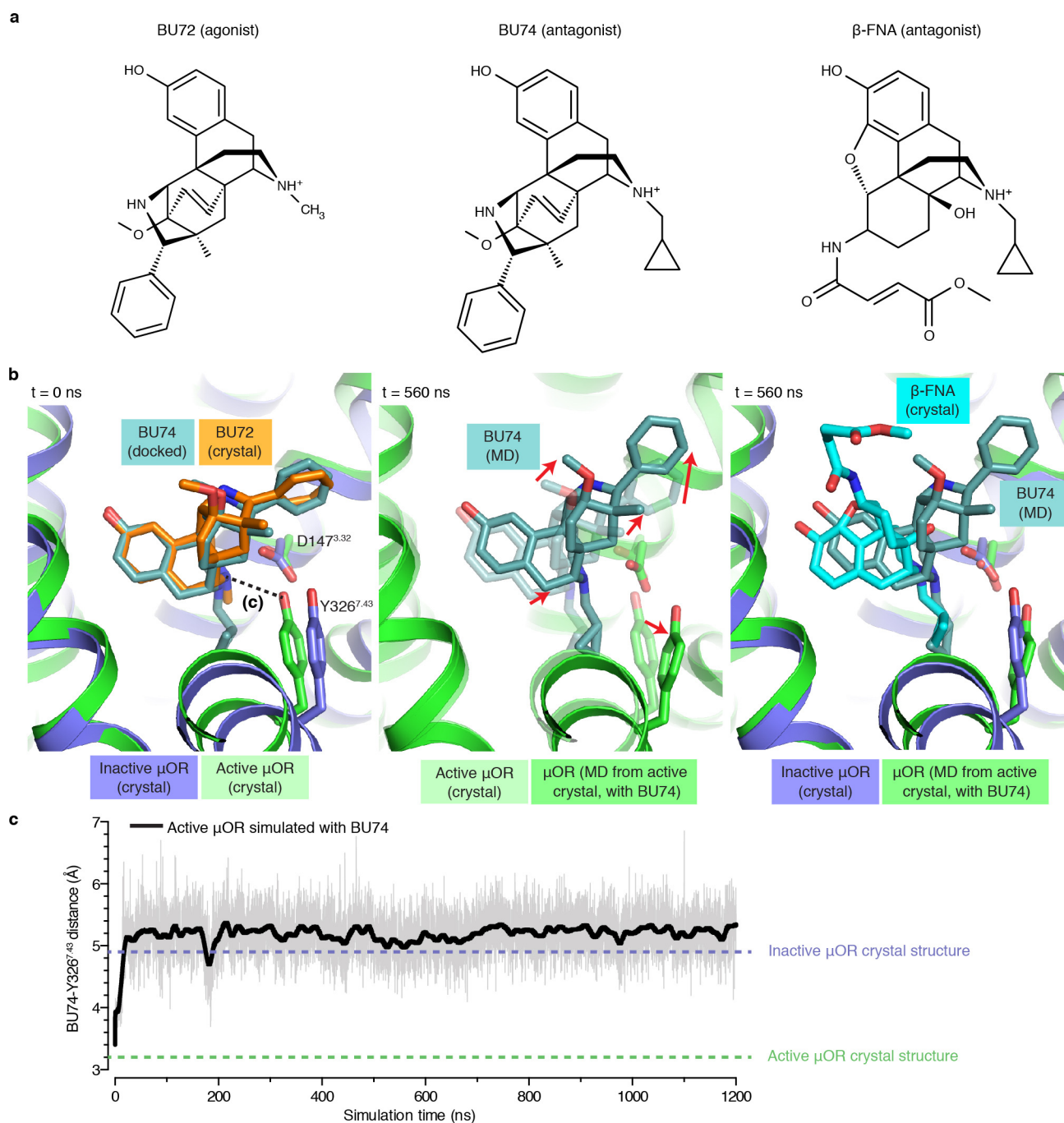
The $2F_o - F_c$ electron density contoured at 2.0σ and within 1.8 Å of residues comprising the active μ OR ligand-binding pocket is shown as grey mesh in **a** and **b**. The same views are shown in **c** and **d** with the omit $F_o - F_c$ density for BU72 displayed as an orange mesh. Displayed $F_o - F_c$ electron density is contoured at 3.0σ . **e**, Placement of an energetically minimized conformation of BU72 within the $F_o - F_c$ electron density shows a poor fit for the pendant phenyl ring. The conformation of BU72 was minimized using quantum mechanical Hartree–Fock methods. **f**, An alternative possible ligand structure with sp^2 geometry at the carbon adjacent to the phenyl (highlighted in red dashed circle) was initially considered due to a better fit within the electron

density. This alternative ligand is predicted to be 2 Da smaller than BU72. **g**, In order to resolve potential ambiguity in the co-crystallized ligand, we performed mass spectrometry on the same protein sample used to generate crystals of the active μ OR. The protein was precipitated in methanol and the supernatant was subjected to MALDI–MS which revealed a strong peak at $m/z = 429.226$, consistent with the expected mass of BU72. **h**, Shown is our final crystallographic model for BU72 within the $F_o - F_c$ electron density. This model probably represents a high-energy conformation of BU72. Notably, the position of the morphinan scaffold is invariant between these alternative models for the crystallized ligand.



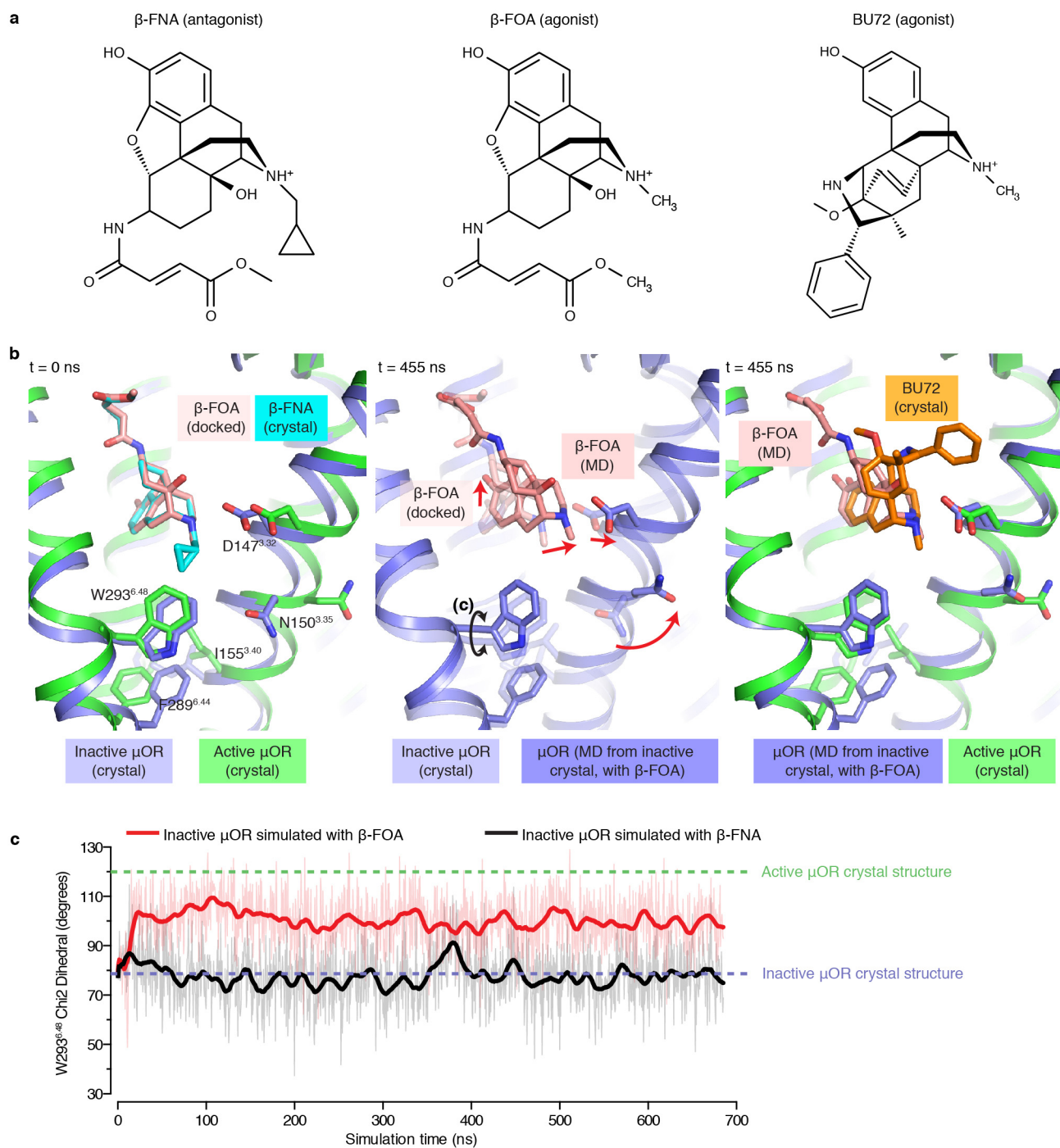
Extended Data Figure 5 | The N terminus of the μ OR interacts with BU72.
a, Surface cut-away view showing that the N terminus forms a lid over the ligand-binding pocket. Shown in the lower panel is the ligand-binding pocket in the absence of the N terminus. **b**, Blue mesh shows the $2F_o - F_c$ omit map

contoured at 1.0σ for the N terminus. **c**, Shown in green mesh is the $F_o - F_c$ omit map contoured at 4.0σ of an unidentified density between BU72 and His54.



Extended Data Figure 6 | Molecular dynamics simulation of active μ OR bound to antagonist BU74. **a**, Structures of agonist BU72, and antagonists BU74 and β -funaltrexamine (β -FNA). The inactive-state structure of μ OR was co-crystallized with β -FNA. **b**, BU74 was docked into the active-state structure of the μ OR based on the crystallographic pose of BU72, but in a molecular dynamics simulation it rapidly moves away from this initial pose. The middle panel highlights the movements of BU74 after 560 ns of simulation and the

rightmost panel shows the comparison of the BU74 pose as compared to the crystal structure of β -FNA bound to inactive μ OR. **c**, Molecular dynamics trajectory measuring the distance between the phenolic hydroxyl of Y326^{7,43} and the tertiary amine of BU74. Dotted lines show the distance between Y326^{7,43} and the same amine of BU72 in the crystal structure of active μ OR and β -FNA in the structure of inactive μ OR.



Extended Data Figure 7 | Molecular dynamics simulation of inactive μ OR bound to agonist β -FOA. **a**, Structures of agonists BU72 and β -fuoxymorphamine (β -FOA) and antagonist β -funaltrexamine (β -FNA). **b**, Molecular dynamics simulation of inactive μ OR with β -FOA docked into the same pose as β -FNA in the inactive-state crystal structure of μ OR. β -FOA shifts towards TM3 with an accompanying rearrangement of TM3

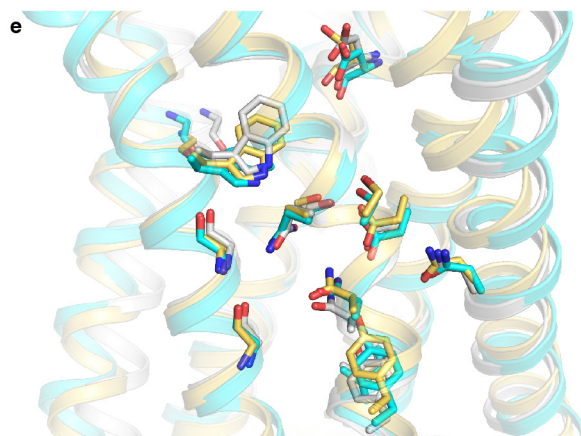
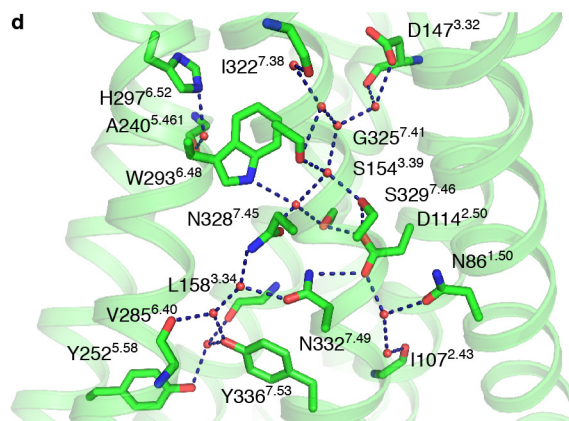
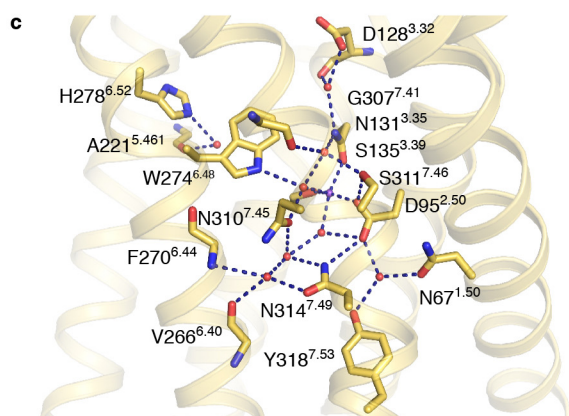
residues D147^{3.32} and N150^{3.35} towards the active-state structure. The overall ligand-binding pocket resembles the active state after 455 ns of simulation. **c**, Trajectory of the W293^{6.48} χ_2 dihedral angle (indicated in the middle panel in **b**) over 700 ns of simulation. In the presence of β -FOA, the preferred rotamer for W293^{6.48} rapidly approaches a conformation similar to the one observed in the structure of active μ OR bound to BU72.

a Polar network residues in inactive states

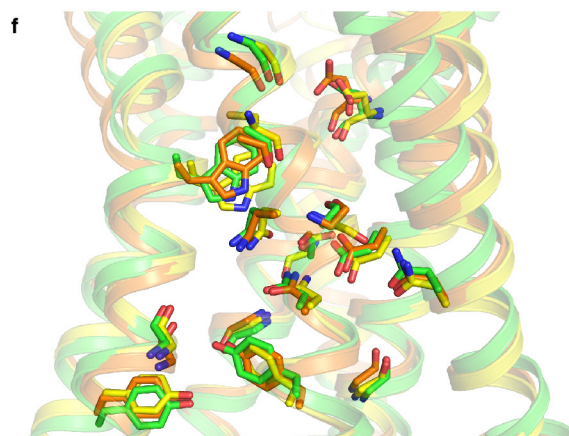
BW numbering	δ OR	β_2 AR	M2R	Rhodopsin
1.50	67N	51N	41N	55N
2.50	95D	79D	96D	83D
3.32	128D	113D	103D	
3.39	135S	120S	110S	
5.46	221A	207S	194A	211H
6.40	266V	278I	392I	257M
6.44	270F	282F	396F	261F
6.48	274W	286W	400W	
7.41	307G	315G	429C	295A
7.45	310N	318N	432N	
7.46	311S	319S	433S	
7.49	314N	322N	436N	302N
7.53	318Y	326Y	440Y	306Y

b Polar network residues in active states

BW numbering	μ OR	β_2 AR	M2R	Rhodopsin
1.50	86N	51N	41N	55N
2.43	107I	72I	62L	76L
2.50	114D	79D	69D	83D
3.32	147D	113D	103D	117A
3.39	154S	120S	110S	
3.43	158L	124L	114L	128L
5.58	252Y	219Y	206Y	223Y
6.40	285V	278I	392I	257M
6.48	293W	286W	400W	265W
7.38	322I	312N	426Y	292A
7.41	325G	315G	429C	295A
7.45	328N	318N	432N	298S
7.46	329S	319S	433S	
7.49	332N	322N	436N	302N
7.53	336Y	326Y	440Y	306Y



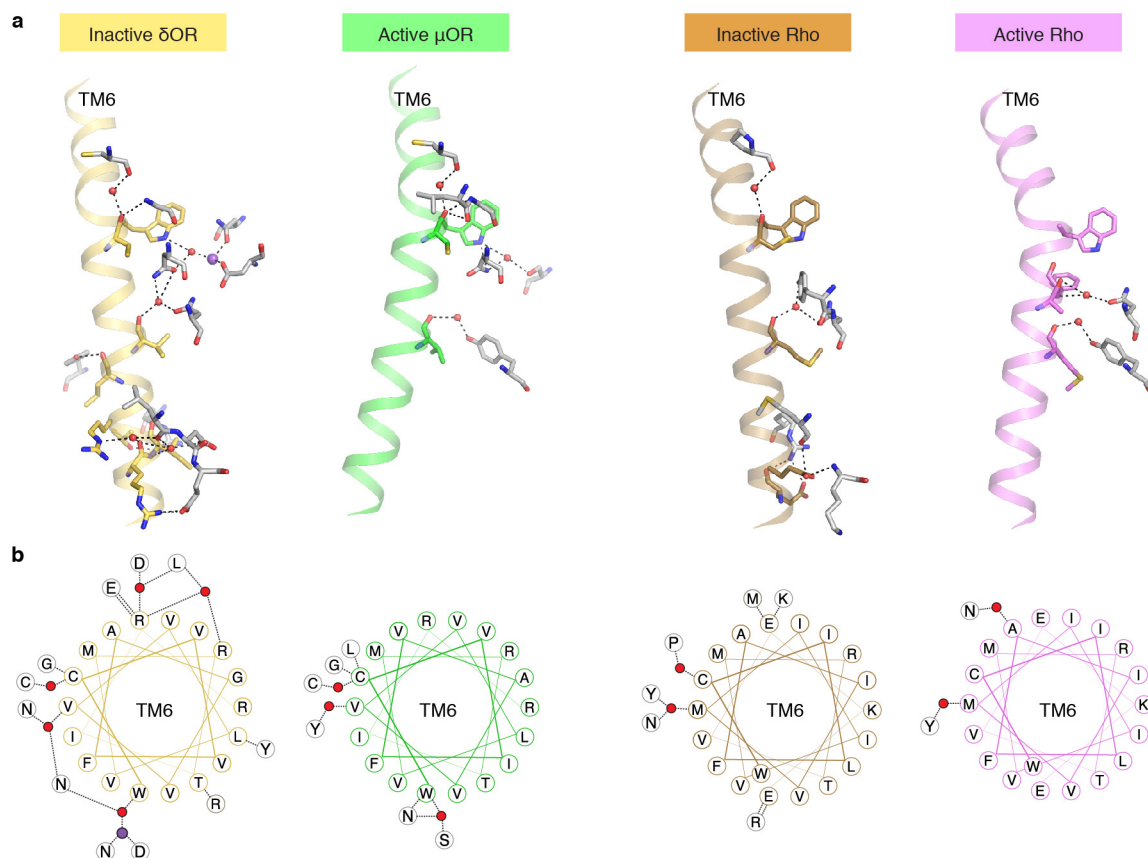
Inactive δ OR Inactive β_2 AR Inactive M2R



Active μ OR Active β_2 AR Active M2R

Extended Data Figure 8 | Comparison of polar networks involved in GPCR activation. **a**, Residues involved in the polar network in the inactive state of the δ OR (PDB ID: 4N6H) and conservation of those residues in β_2 AR, M2R, and rhodopsin. **b**, Residues involved in the polar network in active state μ OR and conservation in β_2 AR, M2R, and rhodopsin. **c**, Water-mediated polar network in the inactive structure of the δ OR involves residues from TM1, TM2,

TM3, TM5, TM6 and TM7. **d**, An identical view as in **c** of the polar network in the active μ OR. **e**, Residues involved in the polar network in inactive structures of δ OR, β_2 AR and M2R are conserved both in sequence and conformation. **f**, In active μ OR, β_2 AR and M2R, the residues within the polar network are again conserved in sequence and conformation.



Extended Data Figure 9 | Differences in TM6 polar network in opioid receptors and rhodopsin. **a**, The entire set of contacts within the polar network that include a residue within TM6 is displayed for the inactive δ OR, active μ OR, and inactive and active rhodopsin (Rho). **b**, Helix wheel representation of

TM6 showing polar contacts. Notably, the inactive δ OR engages in many more polar contacts with neighbouring residues as compared to inactive rhodopsin. Additionally, the active states of both μ OR and rhodopsin have fewer polar contacts than the inactive state.

Extended Data Table 1 | Data collection and refinement statistics (molecular replacement)

	μ OR-BU72-Nb39 ^a
Data collection ^b	
Space group	$I2_12_12_1$
Cell dimensions	
a, b, c (Å)	44.4, 144.0, 209.9
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	50.0-2.07 (2.12-2.07)
R_{merge} (%)	11.2 (153.6)
$\langle I/\sigma I \rangle$	11.0 (1.7)
$CC_{1/2}$ (%)	99.2 (64.7)
Completeness (%)	99.8 (100.0)
Redundancy	8.6 (8.8)
Refinement	
Resolution (Å)	42.5-2.1 (2.15-2.10)
Number of reflections	39,948
$R_{\text{work}}/R_{\text{free}}$ (%)	18.53/22.15 (25.83/28.71)
Number of atoms	
Protein	3,278
Ligand (BU72)	32
Lipid, water and others	208
B-factors (Å ²)	
Protein	56.89
Ligand (BU72)	39.50
Lipid, water and others	69.91
R.M.S. deviation from ideality	
Bond lengths (Å)	0.008
Bond angles (°)	1.193
Ramachandran statistics ^c (%)	
Favored	96.1
Allowed	3.9
Outliers	0

^a Diffraction data from 4 crystals were merged into a complete data set^b Highest resolution shell statistics are shown in parentheses^c As calculated by Molprobit

Growing the gas-giant planets by the gradual accumulation of pebbles

Harold F. Levison¹, Katherine A. Kretke¹ & Martin J. Duncan²

It is widely held that the first step in forming gas-giant planets, such as Jupiter and Saturn, was the production of solid ‘cores’ each with a mass roughly ten times that of the Earth^{1,2}. Getting the cores to form before the solar nebula dissipates (in about one to ten million years; ref. 3) has been a major challenge for planet formation models^{4,5}. Recently models have emerged in which ‘pebbles’ (centimetre-to-metre-sized objects) are first concentrated by aerodynamic drag and then gravitationally collapse to form objects 100 to 1,000 kilometres in size^{6–9}. These ‘planetesimals’ can then efficiently accrete left-over pebbles¹⁰ and directly form the cores of giant planets^{11,12}. This model is known as ‘pebble accretion’; theoretically, it can produce cores of ten Earth masses in only a few thousand years^{11,13}. Unfortunately, full simulations of this process¹³ show that, rather than creating a few such cores, it produces a population of hundreds of Earth-mass objects that are inconsistent with the structure of the Solar System. Here we report that this difficulty can be overcome if pebbles form slowly enough to allow the planetesimals to gravitationally interact with one another. In this situation, the largest planetesimals have time to scatter their smaller siblings out of the disk of pebbles, thereby stifling their growth. Our models show that, for a large and physically reasonable region of parameter space, this typically leads to the formation of one to four gas giants between 5 and 15 astronomical units from the Sun, in agreement with the observed structure of the Solar System.

Our models consist of a series of computer simulations that follow the evolution of a population of objects in a disk around the Sun. The solar composition disk has surface density distribution $\Sigma \propto r_{\text{AU}}^{-1}$, where r_{AU} is the distance from the Sun, and consists of both gas and solids. We assume that an initial population of planetesimals, which follows the surface density of the disk, forms quickly and thus exists at the beginning of our simulations. These planetesimals contain only a small fraction of the mass of solids available (a free parameter of our models). Pebbles either also existed at the beginning of the simulation, or are allowed to form over some period of time (again, a free parameter) starting at the beginning of the calculation. The evolution of this system is followed numerically and includes the effects of gravitational interactions, interactions between bodies in the disk and the gas (although nebular tidal migration is neglected), accretion (including enhancements due to the aerodynamic drag on pebbles; see, for example, ref. 11), and collisional fragmentation. Details of our setup and numerical techniques are described in Methods.

In Fig. 1 we present the results of two different simulations that employ the same parameters except for the method of pebble formation. In the first (Fig. 1a), we assume that the pebbles are left-overs of planetesimal formation and thus they exist at the beginning of the simulation. Note the fast accretion times—embryos (defined to be objects that grow significantly) evolved from having roughly the mass of Pluto to that of Earth in only 1,000 years. However, even though Earth-mass (M_{\oplus}) embryos form quickly, this simulation does not reproduce the Solar System because rather than forming a few cores,

it creates ~ 100 objects each of about $1 M_{\oplus}$. These objects subsequently scatter one another to high-eccentricity, high-inclination orbits, stalling growth. This result is clearly inconsistent with observations. Reference 13 finds that this is a generic outcome for this type of pebble accretion simulation, and thus it could not have occurred in the Solar System. It is indicative of the results one would achieve if the dynamical interactions between planetesimals were neglected (which we refer to as the standard model).

In Fig. 1b, we employ a substantial modification of the pebble accretion theory. We couple an extended timescale for pebble formation (a conjecture supported both by the observed presence of centimetre- to millimetre-sized grains, which are expected to drift rapidly out of protoplanetary disks, in disks of a wide variety of ages¹⁴ and by theoretical models of pebble formation¹⁵) to the dynamical stirring due to planetesimal self-gravity (so-called viscous stirring¹⁶). The behaviour of this simulation is radically different from that in Fig. 1a in that, rather than producing a large population of objects each of $1 M_{\oplus}$, a few giant planets form. Only 5 objects grow to $1 M_{\oplus}$ or larger in this simulation, and there are two gas giants at 10 Myr. Also, note that the timescale for growth is very different in the two calculations. The first object of $1 M_{\oplus}$ grows in 1,000 years in the standard pebble accretion run, as opposed to taking slightly over 400,000 years in this simulation. In the latter case, the growth time is determined by the rate at which pebbles were created.

The outcomes of the two simulations are remarkably different because in the Fig. 1b simulation only the most massive embryos are able to accrete a significant amount of pebbles. This is due to the fact that the embryos grow slowly enough in this model that they can gravitationally interact with one another as they accrete the pebbles. This viscous stirring has two effects on the accretion rates of the embryos. First, it increases the relative velocities between the embryos and the pebbles and thus decreases the capture cross-section (equation (4) in Methods) of the embryos. Second, and more importantly, encounters between embryos lead to increases in the inclinations, i , of the orbits of the embryos and thus the distances they travel above and below the disk midplane. Once the inclinations of the embryos become larger than that of the aerodynamically damped pebbles, the embryos spend much of their time above or below the location where the bulk of the pebbles lie. This effectively starves them and stifles their growth. Owing to the role of viscous stirring in determining the outcome in these simulations, we refer to this process as viscously stirred pebble accretion.

Figure 2 illustrates this effect. In the standard pebble accretion model (Fig. 2a), the system inclinations remain small and thus all the planetesimals can grow. This is because the timescale for viscous stirring is longer than that of the growth. In this case it can be shown (see Methods) that the growth rate of an embryo of mass M_e can be described by $dM_e/dt \propto M_e^{\beta}$ with $\beta \ll 1$, where t is the time since the start of the simulation. The small value of β implies that smaller embryos can catch up with larger ones, leading to a population of like-sized objects.

¹Southwest Research Institute and NASA Solar System Exploration Research Virtual Institute, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA. ²Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, Ontario K7L 3N6, Canada.

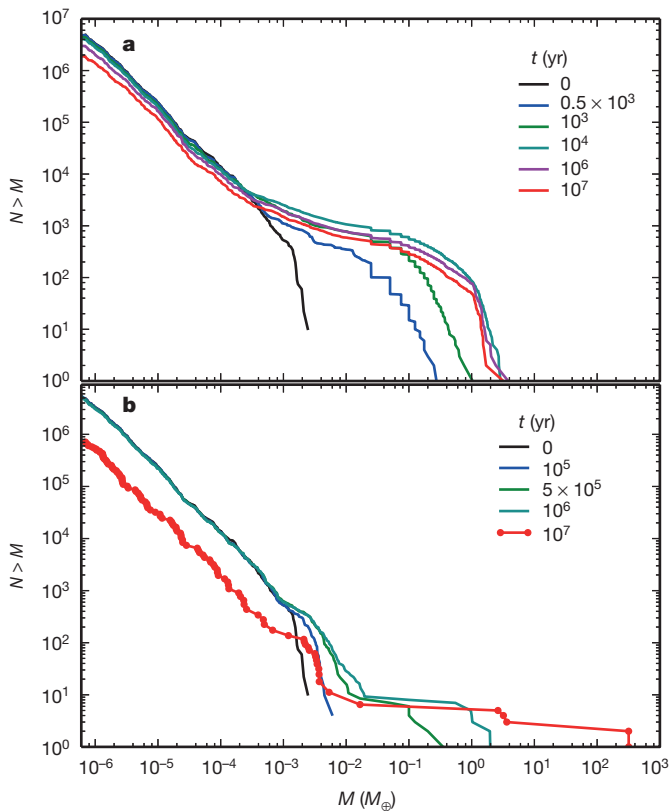


Figure 1 | The cumulative mass distribution of planetesimals and embryos. The growth of planetary embryos (of mass M) in our simulation as illustrated by the number of objects (N) greater than a given mass shown at various times (indicated by colour; see key at right). **a**, An example of standard pebble accretion—all pebbles were in existence at the beginning of the simulation. This is similar to ref. 13. **b**, An example with the same disk parameters as **a** but where the pebbles formed over the lifetime of the disk. This system formed two gas giants and three icy planets with masses between $2 M_{\oplus}$ and $4 M_{\oplus}$. The icy planets are on crossing orbits at the end of the simulation (10 Myr) and thus their number and masses would be likely to change were the calculation run to completion. A video version of this figure is available (Supplementary Video 1).

However, the dynamical evolution of the system is very different in the viscously stirred pebble accretion simulation. Here viscous stirring can act, thereby leading to an increase of inclinations. The magnitude of this increase is set by the mass of the largest embryo—note the increase with time of both the inclination of the smaller embryos and the largest embryo mass in Fig. 2b. Initially the inclinations of the embryos are smaller than those of the pebbles, but by 3,000 years, most of the smaller embryos begin to spend a significant amount of time above or below the pebbles.

The time at which embryos are excited out of the swarm of pebbles can be estimated with a simple calculation. The inclinations of the pebbles are set by the balance between turbulent excitation and aerodynamic drag in the disk. For our example, the average pebble inclination is 0.0016 rad. Given enough time, a population of embryos will stir one another to the point where their relative velocities (v_{rel}) are comparable to their surface escape velocities. Since inclination $i \approx v_{\text{rel}}/v_c$, where v_c is the circular velocity, even if we assume objects do not grow, our population of planetesimals should reach inclinations of ~ 0.06 rad—significantly larger than that of the pebbles. As a result, this effect should be important during much of our simulation. Indeed, Fig. 2b shows that most of the embryos in the system are in this state after only $\sim 3,000$ years.

The most massive embryos, however, behave differently. They tend to have low inclinations (Fig. 2b) as a result of a combination of gravitational interactions with smaller planetesimals (so-called

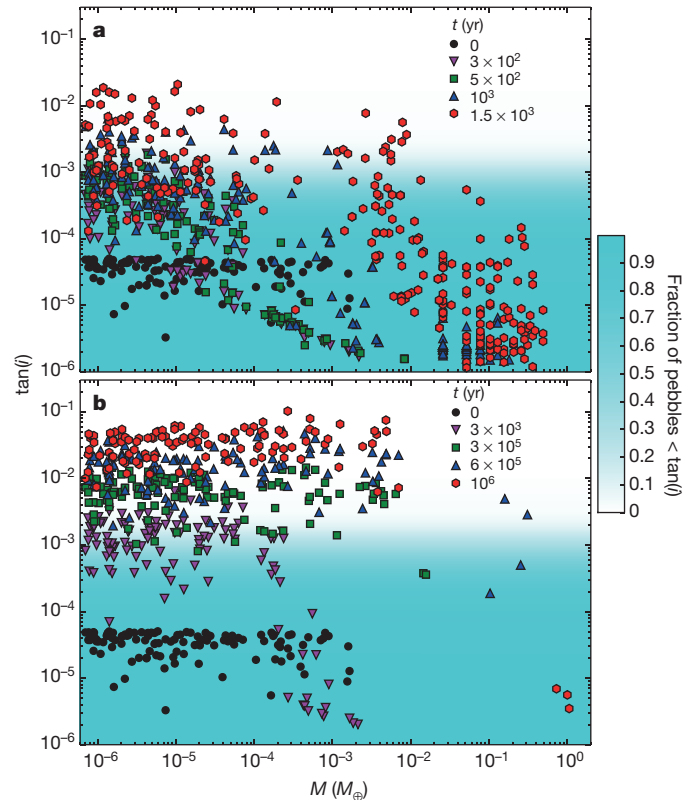


Figure 2 | The vertical distribution of pebbles and embryos. A comparison between the vertical distribution, as represented by $\tan(i)$, of pebbles and embryos in the simulations shown in Fig. 1. The embryos are shown as dots in the figure, where their colour and shape indicates the time within the simulation (see key at top right). Since objects grow at different times within the disk, we only show embryos from a narrow annulus (11.5–13.5 AU). The time averaged cumulative inclination distribution of the pebbles is shown by the cyan gradient (colour key at far right). **a**, The run shown in Fig. 1a; **b**, our fiducial viscously stirred pebble accretion simulation from Fig. 1b. A video version of this figure is available (Supplementary Video 2).

dynamical friction^{16,17}) and with the gas (so-called type I inclination damping¹⁸). As a result, the larger embryos grow relatively quickly while the smaller embryos grow very slowly, if at all. Recall that the standard model of pebble accretion formed a large number of planets because $\beta < 1$, which allowed smaller embryos to catch up with larger ones. Figure 3 shows the relationship between dM_e/dt and M_e in our fiducial viscously stirred pebble accretion run as the system evolves. At early times, β is slightly larger than 1, leading to a few embryos becoming dominant in the population. When the largest objects in the system have masses between $\sim 8 \times 10^{-3} M_{\oplus}$ and $\sim 0.02 M_{\oplus}$, their equilibrium inclinations decrease, leading to a spurt of growth where β is large (~ 4). This allows a small number of embryos to become separated in mass from the rest, and explains why only a small number of embryos become massive enough to become giant-planet cores. Indeed, for the small embryos, dM_e/dt decreases with time because of their increasing inclinations (see Fig. 2). But for $M_e \gtrsim 0.02 M_{\oplus}$, β becomes small again as the strongly damped, large embryos accrete in the same manner as in the standard pebble accretion scenario. This last phase is important because it forces these proto-cores to have similar masses as they grow. Thus, they all reach the mass where they can directly accrete gas at roughly the same time. This last phase might also explain why the four cores of the giant planets probably originally had similar masses. Two gas giants grow in our fiducial simulation (perhaps Jupiter and Saturn). By varying parameters in the models, our systems produced between 0 and 4 gas giants. Results of the simulations can be found in Extended Data Tables 1 and 2.

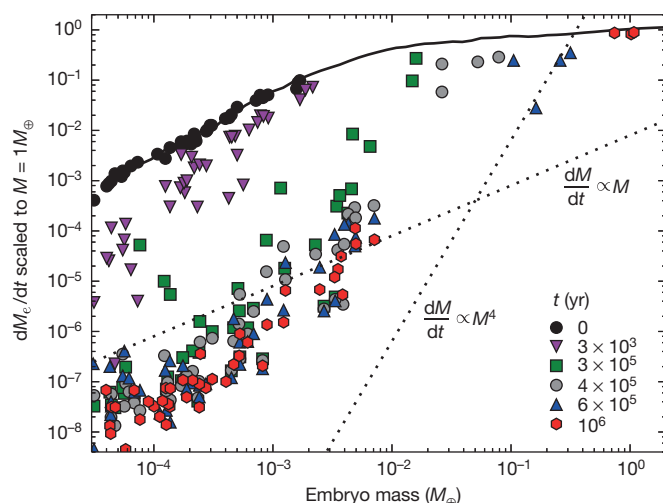


Figure 3 | Embryo growth rate as a function of mass. The temporal evolution of the relationship between embryo growth rate (dM_e/dt) and mass (M_e). The dots show our fiducial viscously stirred pebble accretion simulation (Fig. 1b); their colour and shape indicate the time within the simulation (see key at bottom right). The solid curve is from the standard pebble accretion run in Fig. 1a. These values were calculated using procedures described in Methods. The behaviour of the system is determined by the slope of the data in the figure, β (note that this is a log–log plot), at each time. For reference, the dotted lines show $\beta = 1$ and $\beta = 4$. For $\beta < 1$, small embryos can catch up with larger ones leading to a population of like-sized objects. For $\beta > 1$ the largest embryos run away from their smaller siblings. See Methods for more details.

There is another constraint that any model of giant-planet formation in the Solar System must satisfy in order to be considered a success. The distribution of small bodies in the outer Solar System indicates that the orbits of the giant planets moved substantially after they formed^{19–21}. In particular, Uranus and Neptune are likely to have formed within 15 or 20 AU of the Sun and to have been delivered to their current orbits by either a smooth migration²⁰ or a mild gravitational instability²¹. Both processes require that a population of planetesimals existed on low-eccentricity orbits beyond the giant planets after the planets finished forming. This population must have (1) not formed planets, and (2) survived the planet formation process relatively unscathed. To evaluate this constraint in our simulations, we placed 5 planetesimals with radii similar to Pluto ($s = 1,350$ km) on circular, co-planar orbits with semi-major axes between 20 and 30 AU. None of these objects grew in our fiducial simulation, and they all survived on orbits with eccentricities less than 0.07 (see Methods for how this behaviour varies with disk parameters). Therefore, the process of viscously stirred pebble accretion reproduces the observed structure of the outer Solar System: two gas giants, a few icy planets, and a disk of planetesimals into which the ice giants can migrate.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 October 2014; accepted 9 June 2015.

- Mizuno, H., Nakazawa, K. & Hayashi, C. Instability of a gaseous envelope surrounding a planetary core and formation of giant planets. *Prog. Theor. Phys.* **60**, 699–710 (1978).
- Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
- Haisch, K. E., Lada, E. A. & Lada, C. J. Disk frequencies and lifetimes in young clusters. *Astrophys. J.* **553**, L153–L156 (2001).
- Goldreich, P., Lithwick, Y. & Sari, R. Final stages of planet formation. *Astrophys. J.* **614**, 497–507 (2004).
- Levison, H. F., Thommes, E. & Duncan, M. J. Modeling the formation of giant planet cores. I. Evaluating key processes. *Astron. J.* **139**, 1297–1314 (2010).
- Cuzzi, J. N., Hogan, R. C., Paque, J. M. & Dobrovolskis, A. R. Size-selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophys. J.* **546**, 496–508 (2001).
- Youdin, A. N. & Goodman, J. Streaming instabilities in protoplanetary disks. *Astrophys. J.* **620**, 459–469 (2005).
- Johansen, A. *et al.* Rapid planetesimal formation in turbulent circumstellar disks. *Nature* **448**, 1022–1025 (2007).
- Youdin, A. N. On the formation of planetesimals via secular gravitational instabilities with turbulent stirring. *Astrophys. J.* **731**, 99 (2011).
- Ormel, C. W. & Klahr, H. H. The effect of gas drag on the growth of protoplanets. Analytical expressions for the accretion of small bodies in laminar disks. *Astron. Astrophys.* **520**, A43 (2010).
- Lambrechts, M. & Johansen, A. Rapid growth of gas-giant cores by pebble accretion. *Astron. Astrophys.* **544**, A32 (2012).
- Lambrechts, M. & Johansen, A. Forming the cores of giant planets from the radial pebble flux in protoplanetary discs. *Astron. Astrophys.* **572**, A107 (2014).
- Kretke, K. A. & Levison, H. F. Challenges in forming the Solar System's giant planet cores via pebble accretion. *Astron. J.* **148**, 109 (2014).
- Ricci, L. *et al.* Dust properties of protoplanetary disks in the Taurus-Auriga star forming region from millimeter wavelengths. *Astron. Astrophys.* **512**, A15 (2010).
- Birnstiel, T., Klahr, H. & Ercolano, B. A simple model for the evolution of the dust population in protoplanetary disks. *Astron. Astrophys.* **539**, A148 (2012).
- Stewart, G. R. & Wetherill, G. W. Evolution of planetesimal velocities. *Icarus* **74**, 542–553 (1988).
- Ida, S. Stirring and dynamical friction rates of planetesimals in the solar gravitational field. *Icarus* **88**, 129–145 (1990).
- Ward, W. R. Protoplanet migration by nebula tides. *Icarus* **126**, 261–281 (1997).
- Fernández, J. A. & Ip, W.-H. Some dynamical aspects of the accretion of Uranus and Neptune—the exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
- Malhotra, R. The origin of Pluto's orbit: implications for the Solar System beyond Neptune. *Astron. J.* **110**, 420–429 (1995).
- Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435**, 459–461 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by an NSF Astronomy and Astrophysics Research Grant (principal investigator H.F.L.). We thank A. Johansen, M. Lambrechts, A. Morbidelli, D. Nesvorný and C. Ormel for discussions.

Author Contributions H.F.L. and K.A.K. jointly conceived of the paper and carried out the bulk of the numerical and semi-analytic calculations. M.J.D. developed a semi-analytic model of viscous stirring and growth rates in a population distribution. All authors contributed to the discussion of the results and to the crafting of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.F.L. (hal@boulder.swri.edu).

METHODS

Computational method. For this work we have employed a particle-based Lagrangian code (known as LIPAD, Lagrangian Integrator for Planetary Accretion and Dynamics) that can follow the dynamical/collisional/accretional evolution of a large number of planetesimals through the entire growth process to become planets. For full details about the code and extensive test suites, see ref. 22. For details about how pebble accretion is implemented in the code see ref. 13, but we summarize the most relevant attributes of the code here.

LIPAD is built on top of the N -body integrator SyMBA²³. In order to handle the very large number of sub-kilometre objects required by many simulations, LIPAD utilizes a concept known as tracer particles. Each tracer represents a large number of small bodies with roughly the same orbit and size, and is characterized by three numbers: the physical radius, the bulk density, and the total mass of the disk particles represented by the tracer. LIPAD employs statistical algorithms that follow the dynamical and collisional interactions between the tracers. When a tracer is determined to have been struck by another tracer, it is assigned a new radius according to the probabilistic outcome of the collision based on a fragmentation law by ref. 24, using the ice parameters $Q_0 = 7 \times 10^7 \text{ erg g}^{-1}$, $B = 2.1 \text{ erg cm}^3 \text{ g}^{-2}$, $a = -0.45$ and $b = 1.19$. This way, the conglomeration of tracers represents the size distribution of the evolving planetesimal population. In this work, we do not allow our pebbles to collisionally grow or fragment, therefore particles below 1 km in size are not involved in the collisional cascade.

LIPAD also includes statistical algorithms for viscous stirring, dynamical friction, and collisional damping among the tracers. The tracers mainly dynamically interact with the larger planetary mass objects via the normal N -body routines, which naturally follow changes in the trajectory of tracers due to the gravitational effects of the planets and vice versa. LIPAD is therefore unique in its ability to accurately handle the mixing and redistribution of material due to gravitational encounters, including gap opening and resonant trapping, while also following the fragmentation and growth of bodies. Thus, it is well suited to following the evolution of a population of embryos, planetesimals, and pebbles while they gravitationally interact to form planets.

The pebble accretion model follows the prescription in ref. 10 and is described in detail in ref. 13. An overview of the physics follows in the next section.

In the simulations presented here we do not allow the growing planets to migrate via type I migration¹⁸, although we do include type I eccentricity damping²⁵. We also neglect type II migration. We calculate the aerodynamic drag on all bodies using the formalism of ref. 26. Additionally, as we are interested in the gross evolution of a system after the formation of a potential giant-planet core, we have added a simple optional prescription allowing cores to accrete gas, envelopes. In order to accrete gas, the core size must be above a critical value, which depends on the mass accretion rate of solids onto the core. We follow ref. 27 to determine when core masses are above the critical value given their current mass accretion rate (assuming a grey opacity of 0.02 times that of the interstellar medium²). If this criterion is met then we grow the planet of mass M_p on the Kelvin–Helmholtz timescale (t_{KH}) by accreting gas at a rate of \dot{M}_g , so that:

$$\dot{M}_g = \frac{M_p}{t_{KH}} \quad (1)$$

Following ref. 28 we approximate this timescale as

$$t_{KH} = 10^9 \left(\frac{M_p}{M_\oplus} \right)^{-3} \text{ yr} \quad (2)$$

We limit gas accretion to the Bondi accretion rate,

$$\dot{M}_{g,\text{max}} = \frac{4\pi\rho_g G^2 M_p^2}{c_s^3} \quad (3)$$

where ρ_g is the gas density and c_s is the local sound speed. We note that a major uncertainty in these models is the envelope opacity, which is dominated by the poorly constrained properties of the dust. Varying the opacity dramatically alters both the size of the critical core mass, and perhaps more strikingly, the Kelvin–Helmholtz time. We make two major simplifications in our procedures. First, we arbitrarily cut off gas accretion when a planet reaches one Jupiter mass instead of including any physics to reduce the accretion of gas after gap opening (see, for example, ref. 29). Furthermore, we do not include the fact that pebbles will cease accreting onto planetesimal cores once the cores begin perturbing the disk³⁰. This effect may prove important in causing some cores to accrete gas sooner, or in allowing cores to grow to larger sizes without accreting gas (creating true Uranus and Neptune analogues). Therefore the end masses of giant planets should be viewed with these caveats in mind.

We also note that while this model includes fragmentation, we have found that the production, and subsequent sweep-up of pebbles (as considered by ref. 31) is relatively unimportant in these scenarios. This is due to the assumed relatively low

initial mass of planetesimals in our simulations ($17 M_\oplus$ between 4 AU and 15 AU in our fiducial simulation).

Pebble accretion. Here we present an argument, based on refs 10, 11, about why pebbles are effectively accreted by growing planets. If the stopping time (t_s) of a pebble is comparable to the time it takes to encounter a growing embryo, then it can be deflected out of the gas stream and lose enough orbital energy to become gravitationally bound to the embryo. After capture, the pebble spirals inward due to aerodynamic drag and is accreted. In this case the collisional cross-section for accretion is

$$\sigma_{\text{peb}} \equiv \pi \left(\frac{4GM_e t_s}{v_{\text{rel}}} \right)^2 \exp \left[-2 \left(\frac{t_s v_{\text{rel}}^3}{4GM_e} \right)^\gamma \right] \quad (4)$$

where M_e is the mass of the embryo, v_{rel} is the relative velocity between the pebble and embryo, G is the gravitational constant, and $\gamma = 0.65$ (ref. 10). When the Stokes number of a pebble ($\tau \equiv t_s \Omega_K$, where Ω_K is the local Keplerian frequency) is near unity, this capture radius can be 10^7 times larger than the physical cross-section alone (πR_e^2 , where R_e is the radius of the embryo) for Earth-sized planets in the region inhabited by the giant planets.

This process can cause an isolated 1,000 km object to grow into a $10 M_\oplus$ core in only a few thousand years^{11,13}. To zeroth order, the accretion rate of an embryo growing by pebble accretion is $dM_e/dt = f_{\text{ac}} \mathcal{R}$, where f_{ac} is the fraction of pebbles that drift into an embryo's orbit that will be accreted (also called the filtering factor, which is a function of σ_{peb} and the spatial distribution of pebbles), $\mathcal{R} = 2\pi r \Sigma_{\text{peb}} v_{\text{rad}}$ is the rate at which pebbles are fed to the embryo, Σ_{peb} is the surface density of pebbles at heliocentric distance r , and v_{rad} is the radial velocity of the pebbles due to aerodynamic drag. Pebbles drift at a velocity

$$v_{\text{rad}} = -2 \frac{\tau}{\tau^2 + 1} \eta v_c \quad (5)$$

where η is a dimensionless parameter related to the pressure gradient of the gas disk³² and is of the order of a few times 10^{-3} , and v_c is the local circular velocity of the embryo.

For $\tau \approx 1$, a pebble can spiral through the disk in only a few hundred years. These large values of v_{rad} can lead to huge accretion rates and giant-planet cores can potentially grow quickly.

Pebble formation model. In this work we utilize a simple prescription to convert dust into pebbles over time. We assume that the gas disk with mass M_g exponentially decays over a timescale t_g , so that

$$\dot{M}_g = -\frac{M_g}{t_g} \quad (6)$$

Pebbles (with total mass M_{peb}) are formed at a rate proportional to the square of the dust mass (M_d) such that

$$\dot{M}_{\text{peb}} = k M_d^2 \quad (7)$$

where k determines the rate of pebble formation. The dust in the disk will be lost as the gas disk evolves and as pebbles form, yielding a dust evolution of

$$\dot{M}_d = \dot{M}_g \frac{M_d}{M_g} - \dot{M}_{\text{peb}} \quad (8)$$

This leads to a total rate of pebble production of

$$\dot{M}_{\text{peb}} = \frac{\kappa}{t_g} M_{d,0} \frac{1}{(1+\kappa)^2} \left(\frac{1}{\exp(t/t_g) - \kappa/(1+\kappa)} \right)^2 \quad (9)$$

where $\kappa \equiv k t_g M_{d,0}$. Motivated by observations of disks, we take the time constant for pebble production to be large, with a median production timescale near 1 Myr. We assume that all of the pebbles are produced in 3 Myr. For simplicity, we assume that pebbles are randomly created throughout the disk according to the surface density, and the size of the pebbles is determined by the assumed τ , which is constant in each simulation, but varies between calculations.

To test the robustness of our result to the specific assumptions of our pebble formation model, we modified it in two different ways for the runs shown in Extended Data Tables 1 and 2. In some runs we allow the pebbles to grow as they drift inwards by sweeping up dust. Their growth rate is thus

$$\frac{dm}{dt} = \rho_d \pi r_{\text{peb}}^2 v_{\text{rel}} \quad (10)$$

where ρ_d is the dust mass (which is depleted as pebbles are formed, as dust is swept up, and as the gaseous disk evolves), r_{peb} is the size of the pebble, and v_{rel} is the relative velocity between the pebble and the dust (which is assumed to be perfectly coupled to the gas).

In other runs, we assume a model of pebble formation consistent with the inside out manner suggested by ref. 15. We used the implementation as described in

ref. 12 and had a wave of pebble formation which moves outwards as a function of time. In particular, the radius at which pebbles are generated (r_g) at time t is

$$r_g(t) = \left(\frac{3}{16}\right)^{1/3} (GM_*)^{1/3} (\epsilon_d Z_0)^{2/3} t^{2/3} \quad (11)$$

where M_* is the mass of the Sun, Z_0 is the metallicity of the disk and ϵ_d encapsulates the efficiency of particle growth. We modify ϵ_d to adjust the timescale of pebble formation. We found that so long as the coagulation coefficient is small enough that the timescale for pebble formation is long, we can get results that are generally similar to our fiducial model in which pebbles form randomly.

Model parameters and simulation statistics. Since we are interested in a simulation building the gas-giant planets, we study the growth of planetesimals spread from 4 to 15 AU. As we integrate the simulation forwards, pebbles are formed between 4 and 30 AU. Each system was evolved for 10 Myr using LIPAD. Extended Data Tables 1 and 2 list the 42 simulations that we have completed in our investigation of giant-planet formation. The eight important parameters that were varied are as follows.

- (1) The surface density of the gas disk at 1 AU, Σ_0 . In all our simulations we used a surface density distribution of $\Sigma = \Sigma_0 r_{AU}^{-1}$ (ref. 33), where r_{AU} is the heliocentric distance in AU. For the fiducial simulation, $\Sigma_0 = 7,200 \text{ g cm}^{-2}$, which is 4 times the surface density of the minimum mass solar nebula at that location³⁴. The gas surface density decreases exponentially with a timescale of 2 Myr (refs 3, 35).
- (2) We employ a flaring gas disk with a scale height $h = 0.047 r_{AU}^{3/4}$ AU. We used two values for q : 1.25 from ref. 34 and 9/7 from ref. 36. The latter was used in our fiducial run.
- (3) The size of the largest planetesimal, s_{max} . We draw the initial planetesimals from a distribution of radii, s , of the form $dN/ds \propto s^{-4.5}$ such that s is between 100 km (ref. 37) and s_{max} . For our fiducial simulation $s_{\text{max}} = 1,350 \text{ km}$, which is slightly larger than Pluto.
- (4) The fraction of solids in the disk that are initially converted into planetesimals, f_{pl} . For all simulations, we assume a solid-to-gas ratio of 0.01, which includes the contribution of water ice³⁸, and f_{pl} of the solids are in planetesimals while $(1 - f_{\text{pl}})$ are in pebbles. For the fiducial simulation, $f_{\text{pl}} = 0.1$, which, if extended to 30 AU, is consistent with the mass of planetesimals needed to subsequently deliver the giant planets to their current orbits²¹.
- (5) The initial Stokes number of a pebble, τ . Note that as pebbles spiral towards the Sun, their Stokes number changes (generally decreases) even when it is assumed that they do not grow. For our fiducial simulation $\tau = 0.6$.
- (6) The strength of inclination and eccentricity damping due to the gravitational interaction with the gas disk, c_e . We employ the techniques in ref. 5, which are based on ref. 25, and allow for both radial migration and inclination/eccentricity damping. The strength of both processes can be adjusted by varying two dimensionless parameters: c_a and c_e , respectively. All our simulations have $c_a = 0$, meaning there is no type I migration. For our fiducial simulation $c_e = 0.66$.
- (7) LIPAD²² has the option of allowing pebbles to grow by the accretion of dust particles suspended in the gas. The growth rates are calculated using the particle-in-a-box approximation assuming the distribution of dust follows that of the gas, as described above. Our fiducial simulation has this option disabled.
- (8) The median time of pebble generation, t_{gen} . For most of our simulations, pebble generation follows the evolution of the gas disk. In this case $t_{\text{gen}} \approx 0.7 \text{ Myr}$. However, in two cases we decreased t_{gen} in order to determine whether it affects our results. We found that the value of t_{gen} is not important as long as it is larger than the viscous stirring timescale of the embryos. Times in Extended Data Tables 1 and 2 marked with an asterisk indicate that pebbles were generated in an inside-out manner as described above.

Additionally, all the disks are turbulent with $\alpha = 4 \times 10^{-3}$ (ref. 39).

The last five columns in Extended Data Tables 1 and 2 present the basic characteristics of our systems. In particular, N_{Gas} is the total number of gas-giant planets that formed during each run. It ranges from 0 to 4. Occasionally, a giant planet or two was lost from the system. Although some of these were ejected owing to dynamical instabilities, the majority were pushed off the inner edge of our computational domain by their neighbours as the latter accreted gas. Unfortunately, the inner edge of the domain needs to be relatively large because it determines the timestep used in the calculation. The N -body integrator requires that the perihelion passage be temporally resolved, and since the location of the domain's inner boundary determines the smallest perihelion, it also sets the timestep. In order for the calculations to be practical, we were forced to remove any object with a perihelion distance smaller than 2.7 AU (the location of the snow line). The majority of the lost gas giants would probably have remained in the asteroid belt or migrated into the terrestrial planet region if we had been able to keep them in the calculations. The column labelled $N_{\text{f}}^{\text{Gas}}$ gives the number of gas giants remaining in the system at 10 Myr.

Columns (11) and (12) in Extended Data Tables 1 and 2 refer to icy objects with masses greater than $1 M_{\oplus}$. Again, $N_{\text{ice}}^{\text{ice}}$ and $N_{\text{ice}}^{\text{ice}}$ are the total number of planets that

formed and the number remaining at 10 Myr, respectively. Unlike their larger siblings, the majority of the lost icy planets were ejected from the planetary system, mainly by encounters with growing gas giants. The number of lost ice planets illustrates the fact that the system becomes violent after the largest cores begin to accrete gas. The last column shows whether the 5 planetesimals we placed on circular co-planar orbits between 20 and 30 AU (as proxies for the planetesimal disk needed for later planet migration²¹) survived at 10 Myr.

There are a couple of important caveats that must be taken into account to properly interpret the results in Extended Data Tables 1 and 2. First, the parameter space for our calculations is 8-dimensional, and each calculation required many weeks of CPU time. This limited the total number that could be performed. Thus, it was impossible to cover parameter space in a meaningful way. Instead, we surgically approached the problem by choosing a couple of starting locations and varying individual parameters to test their effects. We then moved in a particular direction in parameter space if we believed we would be more likely to produce systems with some desired characteristic. As a result, the distribution of our results seen in Extended Data Tables 1 and 2 (for example the number of gas giants) cannot be interpreted as the true distribution that would result if parameter space were uniformly covered.

Our second caveat is the crude methods we used to model the direct accretion of gas onto the cores to form gas giants. As discussed above, our simple model for calculating the onset of gas accretion and the final mass of planets is probably missing important physics. These limitations mean that, although our calculations show that pebble accretion produces the correct number of giant-planet cores on the correct timescales, the details of the evolution of our systems after the gas giants form should be viewed with scepticism.

Generating Fig. 3. The data in Fig. 3 were generated using the following procedure. We first needed to construct a high-resolution distribution of pebbles. This was accomplished by summing the output steps of our integration over all time. Then at each time plotted, we noted the mass and orbit of each embryo. We then generated 1,000 clones of each embryo—each with the same semi-major axis, eccentricity, and inclination, but with different orbital angles. For each clone, we calculated its velocity with respect to the pebbles and then σ_{peb} from equation (4). From this, we estimate the instantaneous accretion rate using the particle-in-a-box approximation, $\dot{M}_c = \rho \sigma_{\text{peb}} v_{\text{rel}}$ where ρ is the local mass density of pebbles. The dots in the figure show the average of the 1,000 clones of the instantaneous accretion rates for each embryo.

Code availability. The calculations presented in this paper were performed using LIPAD, a proprietary software product funded by the Southwest Research Institute that is not publicly available. It is based upon the N -body integrator SyMBA, which is publicly available at <http://www.boulder.swri.edu/swifter/>.

22. Levison, H. F., Duncan, M. J. & Thommes, E. A Lagrangian integrator for planetary accretion and dynamics (LIPAD). *Astron. J.* **144**, 119–138 (2012).
23. Duncan, M. J., Levison, H. F. & Lee, M. H. A multiple time step symplectic algorithm for integrating close encounters. *Astron. J.* **116**, 2067–2077 (1998).
24. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142**, 5–20 (1999).
25. Papaloizou, J. C. B. & Larwood, J. D. On the orbital evolution and growth of proto-planets embedded in a gaseous disc. *Astronomy* **315**, 823–833 (2000).
26. Adachi, I., Hayashi, C. & Nakazawa, K. The gas drag effect on the elliptical motion of a solid body in the primordial solar nebula. *Prog. Theor. Phys.* **56**, 1756–1771 (1976).
27. Rafikov, R. R. Atmospheres of protoplanetary cores: critical mass for nucleated instability. *Astron. J.* **648**, 666–682 (2006).
28. Ida, S. & Lin, D. N. C. Toward a deterministic model of planetary formation. IV. Effects of type I migration. *Astrophys. J.* **673**, 487–501 (2008).
29. Dobbs-Dixon, I., Li, S. L. & Lin, D. N. C. Tidal barrier and the asymptotic mass of proto-gas giant planets. *Astrophys. J.* **660**, 791–806 (2007).
30. Lambrechts, M., Johansen, A. & Morbidelli, A. Separating gas-giant and ice-giant planets by halting pebble accretion. *Astron. Astrophys.* **572**, A35 (2014).
31. Chambers, J. E. Giant planet formation with pebble accretion. *Icarus* **233**, 83–100 (2014).
32. Nakagawa, Y., Sekiya, M. & Hayashi, C. Settling and growth of dust particles in a laminar phase of a low-mass solar nebula. *Icarus* **67**, 375–390 (1986).
33. Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C. & Dullemond, C. P. Protoplanetary disk structures in Ophiuchus. II. Extension to fainter sources. *Astrophys. J.* **723**, 1241–1254 (2010).
34. Hayashi, C. Structure of the solar nebula, growth and decay of magnetic fields and effects of magnetic and turbulent viscosities on the nebula. *Prog. Theor. Phys.* **70** (Supplement), 35–53 (1981).
35. Hernández, J. et al. Spitzer observations of the λ Orionis cluster. I. The frequency of young debris disks at 5 Myr. *Astrophys. J.* **707**, 705–715 (2009).
36. Chiang, E. I. & Goldreich, P. Spectral energy distributions of T Tauri stars with passive circumstellar disks. *Astrophys. J.* **490**, 368–376 (1997).
37. Morbidelli, A., Bottke, W. F., Nesvorný, D. & Levison, H. F. Asteroids were born big. *Icarus* **204**, 558–573 (2009).
38. Lodders, K. Solar System abundances and condensation temperatures of the elements. *Astrophys. J.* **591**, 1220–1247 (2003).
39. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24**, 337–355 (1973).

Extended Data Table 1 | Our completed simulations.

Table 1: Our completed simulations.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Σ_0	q	s_{\max}	f_{pl}	τ	c_e	pebbles grow?	t_{gen}	N^{Gas}	N_f^{Gas}	N^{ice}	N_f^{ice}	Disk ok?
(g/cm ²)		(km)				(Y/N)	(Myr)					(Y/N)
3600	9/7	1500	0.2	0.3	1	N	0.7	4	2	12	0	Y
3600	9/7	1500	0.2	0.3	1	N	0.5	4	3	14	0	Y
3600	9/7	1500	0.2	0.3	1	N	0.2	4	3	15	1	N
3600	9/7	1500	0.2	0.4	1	N	0.7	3	1	9	2	Y
3600	1.25	1500	0.2	3	1	Y	0.7	3	2	9	0	Y
4050	9/7	1500	0.18	0.4	1	N	0.7	4	2	6	1	Y
4050	9/7	1500	0.18	0.5	1	N	0.7	3	3	7	0	Y
4050	9/7	1500	0.18	0.6	1	N	0.7	4	4	9	0	Y
4500	9/7	1500	0.16	0.3	1	N	0.7	4	3	13	0	Y
4500	9/7	1500	0.16	0.4	1	N	0.7	3	2	7	1	Y
4500	9/7	1500	0.16	0.5	1	N	0.7	3	3	13	0	N
4500	9/7	1500	0.16	0.5	0.33	N	0.7	4	2	4	0	Y
4500	9/7	1500	0.16	0.5	0.1	N	0.7	0	0	7	6	Y
4500	9/7	1500	0.16	0.6	1	N	0.7	2	2	15	3	N
4500	9/7	1500	0.16	0.7	1	N	0.7	0	0	7	7	Y
4500	9/7	1500	0.16	0.9	1	N	0.7	1	0	1	1	Y
5400	1.25	1500	0.15	3	1	Y	0.7	0	0	5	4	Y
5400	1.25	1500	0.15	3	1	Y	0.7	0	0	6	6	Y
5400	1.25	1500	0.15	3	0.66	Y	0.7	0	0	5	5	Y
5400	1.25	1500	0.15	3	0.33	Y	0.7	1	1	7	4	Y
7200	9/7	1350	0.25	0.6	1	N	0.7	4	4	10	2	Y
7200	9/7	1350	0.15	0.6	1	N	0.7	4	4	8	0	N
7200	1.25	1500	0.1	1	1	Y	0.7	2	2	11	4	Y
7200	1.25	1500	0.1	1	0.25	Y	0.7	3	3	5	1	Y
7200	1.25	1500	0.1	1	0	Y	0.7	2	2	1	0	Y
7200	9/7	1350	0.1	0.6	1	N	0.7	2	2	8	2	Y
7200	9/7	1350	0.1	0.6	0.66	N	0.7	2	2	9	3	Y
7200	9/7	1350	0.1	0.6	0.66	N	0.7	3	3	5	1	Y
7200	9/7	1350	0.05	0.6	1	N	0.7	2	2	2	1	Y
7200	9/7	1350	0.025	0.6	1	N	0.7	3	3	13	0	N

See Methods for a full description of the columns in this table.

Extended Data Table 2 | More simulations.

Table 2: More Simulations.												
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Σ_0	q	s_{\max}	f_{pl}	τ	c_e	pebbles grow?	t_{gen}	N^{Gas}	N_f^{Gas}	N^{ice}	N_f^{ice}	Disk ok?
(g/cm ²)		(km)				(Y/N)	(Myr)					(Y/N)
900	9/7	1500	0.8	0.1	0.66	N	0.9*	0	0	21	15	Y
900	9/7	1500	0.8	0.1	0.66	N	1.3*	0	0	17	14	Y
1800	9/7	1500	0.4	0.1	0.66	N	0.9*	5	4	8	0	N
1800	9/7	1500	0.4	0.1	0.66	N	1.3*	3	3	6	0	N
3600	9/7	1500	0.2	0.1	0.66	N	0.9*	7	5	12	0	N
3600	9/7	1500	0.2	0.1	0.66	N	1.3*	6	4	9	0	N
3600	9/7	1500	0.2	0.1	0.66	N	0.4*	4	3	9	0	N
3600	9/7	1500	0.2	0.6	0.66	N	0.9*	5	3	6	0	N
3600	9/7	1500	0.2	0.6	0.66	N	0.4*	2	2	11	4	Y
7200	9/7	1500	0.1	0.6	0.66	N	0.4*	3	1	10	1	N
7200	9/7	1500	0.1	0.6	0.66	N	0.9*	4	3	6	0	N
7200	9/7	1500	0.1	0.6	0.66	N	1.3*	5	3	5	1	N

This is a continuation of Extended Data Table 1. See Methods for a full description of the columns in this table.

Measurement-based control of a mechanical oscillator at its thermal decoherence rate

D. J. Wilson¹, V. Sudhir¹, N. Piro¹, R. Schilling¹, A. Ghadimi¹ & T. J. Kippenberg¹

In real-time quantum feedback protocols^{1,2}, the record of a continuous measurement is used to stabilize a desired quantum state. Recent years have seen successful applications of these protocols in a variety of well-isolated micro-systems, including microwave photons³ and superconducting qubits⁴. However, stabilizing the quantum state of a tangibly massive object, such as a mechanical oscillator, remains very challenging: the main obstacle is environmental decoherence, which places stringent requirements on the timescale in which the state must be measured. Here we describe a position sensor that is capable of resolving the zero-point motion of a solid-state, 4.3-megahertz nanomechanical oscillator in the timescale of its thermal decoherence, a basic requirement for real-time (Markovian) quantum feedback control tasks, such as ground-state preparation. The sensor is based on evanescent optomechanical coupling to a high-*Q* microcavity⁵, and achieves an imprecision four orders of magnitude below that at the standard quantum limit for a weak continuous position measurement⁶—a 100-fold improvement over previous reports^{7–9}—while maintaining an imprecision-back-action product that is within a factor of five of the Heisenberg uncertainty limit. As a demonstration of its utility, we use the measurement as an error signal with which to feedback cool the oscillator. Using radiation pressure as an actuator, the oscillator is cold damped¹⁰ with high efficiency: from a cryogenic-bath temperature of 4.4 kelvin to an effective value of 1.1 ± 0.1 millikelvin, corresponding to a mean phonon number of 5.3 ± 0.6 (that is, a ground-state probability of 16 per cent). Our results set a new benchmark for the performance of a linear position sensor, and signal the emergence of mechanical oscillators as practical subjects for measurement-based quantum control.

Feedback control of mechanical oscillation is found in many applications, ranging from stabilizing the approximately-kilogram test mass in a gravity-wave interferometer¹¹ to trapping/cooling of atomic¹² and sub-atomic particles¹³. A basic approach to feedback control uses a sensor to track the position of the oscillator and an actuator to convert the measurement record into a continuous and prompt ('real-time') feedback force. Recently, the quantum limits of continuous feedback² have been explored, enabling demonstrations such as microwave Fock-state generation³ and persistent Rabi oscillations in a superconducting qubit⁴. These protocols employ a 'weak measurement' capable of tracking a quantum state as rapidly as it decoheres due to measurement back-action¹⁴. For mechanical oscillators, ideal weak position measurements⁶ have been available since the advent of the laser, in the context of shot-noise-limited interferometry¹⁵. Only recently, however, with the confluence of low-loss, cryogenic micromechanics and on-chip, integrated photonics (as explored in contemporary cavity optomechanics¹⁶), has it been feasible to consider their application to quantum feedback protocols^{17,18}. The main challenge is thermal noise, which places stringent requirements on the timescale in which the measurement must take place.

Feedback cooling^{10,12,17,19,20} is a well-studied control protocol that illustrates both the utility and the challenge of quantum feedback applied to mechanical systems. In feedback cooling protocols, a mech-

anical oscillator undergoing thermal Brownian motion is steered towards its ground state by minimizing a measurement of its displacement, S_x (here expressed as a power spectral density⁶ evaluated at the mechanical oscillator frequency, Ω_m). A conventional strategy¹⁰ is to apply a feedback force proportional to the time derivative of the measurement (which estimates the velocity of the oscillator). The resultant damping acts to reduce (cool) the motion of the oscillator until it coincides with the measurement imprecision, S_x^{imp} . Cooling the oscillator until it spends most of the time in its ground state (a mean phonon occupancy of $n_m < 1$) is possible if S_x^{imp} remains lower than the zero-point fluctuations of the damped oscillator, that is, if $S_x^{\text{imp}} \lesssim S_x^{\text{zp}}/n_{\text{th}}$ (see Supplementary Information), where S_x^{zp} is the spectral density of intrinsic (absent feedback) zero-point fluctuations and n_{th} is the phonon occupancy of the thermal bath. ($S_x^{\text{zp}} = 4x_{\text{zp}}^2/\Gamma_m$ is proportional to the ratio of the ground-state variance $x_{\text{zp}}^2 = \hbar/(2m\Omega_m)$ and the intrinsic mechanical damping rate Γ_m , where \hbar is the reduced Planck's constant and m is the effective mass of the oscillator.) In the frequency domain, this condition on S_x^{imp} amounts to resolving the intrinsic thermal displacement $S_x \approx 2n_{\text{th}}S_x^{\text{zp}}$ with a signal-to-noise ratio of $S_x/S_x^{\text{imp}} \gtrsim 2n_{\text{th}}^2$; in the time domain, it corresponds to resolving the zero-point motion at a characteristic 'measurement' rate⁶

$$\Gamma_{\text{meas}} \equiv \frac{x_{\text{zp}}^2}{2S_x^{\text{imp}}} \gtrsim \frac{\Gamma_{\text{th}}}{8} \quad (1)$$

where $\Gamma_{\text{th}} \approx \Gamma_m n_{\text{th}}$ is the thermal decoherence rate of the oscillator. Equation (1) implies the ability to resolve a displacement of x_{zp} in the timescale over which a single phonon enters from the thermal bath, Γ_{th}^{-1} . Satisfying this requirement is a technically daunting challenge because of the small x_{zp} and large Γ_{th} of typical engineered mechanical oscillators. As a consequence, despite the success of autonomous feedback^{21–23}, ground-state cooling using measurement-based feedback has yet to be demonstrated.

An additional, fundamental caveat compounds the challenge of ground-state cooling and hints at the underlying virtue of quantum feedback: Heisenberg's uncertainty principle predicts that a weak ($\Gamma_{\text{meas}} \ll \Omega_m$) continuous position measurement with an imprecision of $S_x^{\text{zp}}/2$ will produce a stochastic 'back-action' force that disturbs the position of the oscillator by at least the same amount⁶. By inference, a thermal-phonon-equivalent imprecision of $n_{\text{imp}} \equiv S_x^{\text{imp}}/(2S_x^{\text{zp}})$ results in an effective increase of the thermal bath occupation by $n_{\text{ba}} \geq 1/(16n_{\text{imp}})$ (see Supplementary Information). This penalty would appear to prohibit ground-state cooling, as it entails substantially heating the oscillator to achieve the necessary measurement precision. Remarkably, however, feedback counteracts back-action²⁴, so that a phonon occupancy of $n_m \approx 2\sqrt{n_{\text{imp}}(n_{\text{ba}} + n_{\text{th}})} - 1/2 < 1$ (see Supplementary Information) can still be achieved^{17,20}. The limiting case of $n_m \rightarrow 0$ is approached when the measurement record is dominated by back-action-induced fluctuations. This occurs when the measurement is maximally efficient¹⁴, that is, when the measurement rate, $\Gamma_{\text{meas}} = \Gamma_m/(16n_{\text{imp}})$, approaches the effective thermal decoherence rate, $\Gamma_{\text{tot}} \approx (n_{\text{th}} + n_{\text{ba}})\Gamma_m$. To meet this condition for a typical

¹Institute of Condensed Matter Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

engineered oscillator, a linear position sensor must achieve an imprecision far (about n_{th} times) below the natural scale set by the ‘standard quantum limit’ (SQL)⁶ ($n_{\text{imp}} = n_{\text{ba}} = 1/4$), while maintaining back action near the uncertainty limit: $4\sqrt{n_{\text{imp}}n_{\text{ba}}} \geq 1$.

Coupling micromechanical oscillators to optical cavities has emerged as a promising way to meet the above requirements. Transduction in such ‘cavity-optomechanical’ systems¹⁶ arises from a parametric coupling, $G = \partial\omega/\partial x$, between the position of the oscillator and the resonance frequency ω_c of a cavity. For broadband sensing, characterized by a cavity decay rate $\kappa \gg \Omega_m$, a resonant laser field passing through the cavity acquires a phase shift of $2G\delta x/\kappa$; this can be resolved in a homodyne interferometer with a quantum-noise-limited imprecision of $S_x^{\text{imp}} = (8G^2 n_c \eta / \kappa)^{-1}$, where n_c is the mean intracavity photon number and $\eta \in [0, 1]$ is the effective photon collection efficiency (see Supplementary Information). The associated measurement rate is $\Gamma_{\text{meas}} = 4g_0^2 n_c \eta / \kappa \equiv \Gamma_m C_0 n_c \eta$, where $g_0 \equiv Gx_{\text{zp}}$ is the vacuum optomechanical coupling rate and $C_0 \equiv 4g_0^2 / (\kappa \Gamma_m)$ is the ‘single-photon

cooperativity”¹⁶, which characterizes the per-photon measurement rate. To achieve efficient measurements, contemporary cavity-optomechanical systems combine the state-of-the-art in high-Q nanomechanics and microphotonics¹⁶. As a consequence, imprecision below that at the SQL^{7–9}, as well as quantum back-action (that is, radiation pressure shot noise¹⁵)^{25,26} has recently been observed. In none of these experiments, however, was $\Gamma_{\text{meas}} \approx \Gamma_{\text{th}}$ demonstrated at the photodetector, owing to a combination of large thermal occupation, extraneous imprecision, optical loss and dynamic instabilities. (An electromechanical device operating deep in the ‘good-cavity’ limit, $\kappa \ll \Omega_m$, has recently achieved $\Gamma_{\text{meas}} > \Gamma_{\text{th}}$ (ref. 22); however, the use of a far-off-resonant probe resulted in strong dynamic coupling of the optical and mechanical mode.)

Our system addresses these challenges using near-field optomechanical coupling⁵, a paradigm for combining mechanical and optical resonators with differing material and geometry. Building on earlier work^{5,27}, we couple a mechanical oscillator with an exceptionally

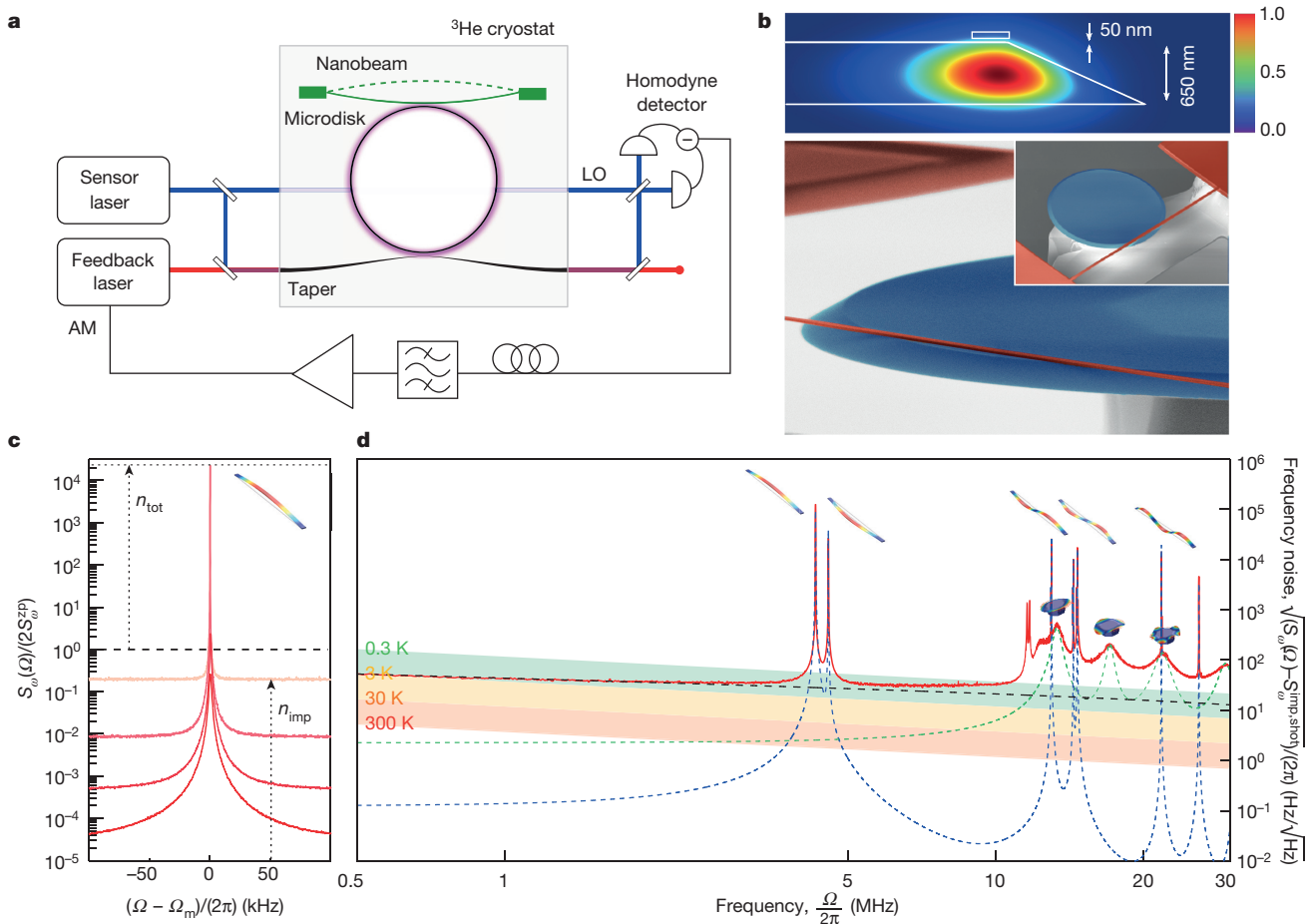


Figure 1 | Measuring and controlling the position of a nanomechanical beam using a near-field optomechanical transducer. **a**, Whispering gallery modes of a SiO₂ microdisk are excited using a tapered optical fibre driven by a pair of tunable diode lasers (sensor and feedback). Displacement of a Si₃N₄ nanobeam, which samples the evanescent mode volume of the microdisk, is recorded in the phase of the transmitted sensor field using a balanced homodyne detector (LO, local oscillator field). Radiation-pressure feedback is applied by modulating the amplitude of the feedback laser with an electronically processed (delayed, bandpass-filtered and amplified) copy of the homodyne photocurrent (AM, amplitude modulation). Sample and fibre are both embedded in a ³He cryostat operating at 4.4 K. **b**, Above, finite element model of the optical mode (colour scale shows normalized field amplitude), shown in vertical cross-section. Wedged disk and beam (separated by 50 nm) are outlined in white. Optomechanical coupling is proportional to the intensity gradient at the position of the beam. Below, scanning electron micrograph of the device. False-colouring indicates material: blue, SiO₂ (microdisk); red, Si₃N₄

(nanobeam); grey, Si (wafer substrate). **c**, Thermal displacement of the fundamental beam mode, expressed as a power spectral density of cavity frequency noise, $S_{\omega}(\Omega)$ (where Ω is the angular Fourier frequency). Various measurement strengths (different n_c) are shown. In units scaled to the zero point spectral density, S_{ω}^p , the peak and background correspond to noise quanta n_{tot} and n_{imp} , respectively (see text). **d**, Broadband (shot-noise-subtracted) homodyne signal expressed as apparent cavity-frequency noise; the spectral density is divided by 2π because ω is an angular frequency. Solid red line corresponds to measurement data; dashed blue, green, and black lines correspond to estimated contributions from nanobeam thermomechanical, microdisk thermomechanical, and microdisk thermorefractive noise, respectively. Coloured bands denote the imprecision required for $\Gamma_{\text{meas}} = \Gamma_{\text{th}}$ at different bath temperatures T , as labelled: $S_{\omega}^{\text{imp}} = g_0^2 \hbar Q_m / (2k_B T)$, assuming $g_0^2 \propto 1/\Omega$ and $Q_m = 7.6 \times 10^5$. Noise peaks arising from microdisk and nanobeam thermal motion are labelled by numerical models of the corresponding mode shapes (not to scale).

high $Q/(\text{mass})$ ratio and low optical absorption—a high-stress Si_3N_4 nanomechanical beam—to an optical cavity with a high $Q/(\text{mode volume})$ ratio and low optical nonlinearity—a SiO_2 microdisk—by localizing a portion of the beam within the evanescent volume of one of the whispering gallery modes of the microdisk (Fig. 1b). The system is integrated onto a silicon chip, allowing for robust cryogenic operation. In contrast to earlier work, we employ a fabrication technique (see Supplementary Information) that substantially reduces the distance between the mechanical and optical element, enabling separation of the Si_3N_4 and SiO_2 surfaces by 10–100 nm. In conjunction with cryogenic operation, we realize an enhancement of C_0 by nearly five orders of magnitude compared to the device reported in ref. 27, while reducing the thermal bath occupation by two orders of magnitude.

We study a system consisting of a $65\text{ }\mu\text{m} \times 400\text{ nm} \times 70\text{ nm}$ (effective mass $m \approx 2.9\text{ pg}$) nanobeam placed approximately 50 nm from the surface of a microdisk with a diameter of $30\text{ }\mu\text{m}$. The microdisk is optically probed using a low-loss (approximately 6%) fibre taper that is pumped by a tunable diode laser. Displacement of the nanobeam is observed in the phase of the transmitted cavity field using a balanced homodyne interferometer. We interrogate two optical modes: a ‘sensor’ mode (used for readout) at $\lambda_c \approx 775\text{ nm}$ that exhibits an intrinsic decay rate of $\kappa_0 \approx 2\pi \times 0.44\text{ GHz}$ and a ‘feedback’ mode (used for radiation pressure actuation) at $\lambda_c \approx 843\text{ nm}$ for which $\kappa_0 \approx 2\pi \times 1\text{ GHz}$. For the mechanical oscillator, we use the $\Omega_m \approx 2\pi \times 4.3\text{ MHz}$ fundamental out-of-plane mode of the nanobeam. The vacuum optomechanical coupling rate between the oscillator and the sensor mode is $g_0 \approx 2\pi \times 20\text{ kHz}$, which corresponds to a ‘frequency pulling factor’ of $G \approx 2\pi \times 0.70\text{ GHz nm}^{-1}$ for the estimated zero-point amplitude of $x_{\text{zp}} \approx 29\text{ fm}$. Our experiments are conducted in a ^3He buffer gas cryostat²³ at a bath temperature of $T \approx 4.4\text{ K}$ ($n_{\text{th}} \approx k_B T / (\hbar \Omega_m) \approx 2.1 \times 10^4$) and at gas pressures below 10^{-3} mbar . Ringdown measurements reveal a mechanical damping rate of $\Gamma_m \approx 2\pi \times 5.7\text{ Hz}$ (corresponding to a quality factor $Q_m = \Omega_m / \Gamma_m \approx 7.6 \times 10^5$). Our system thus allows for a near-unity single-photon cooperativity of $C_0 \approx 0.64$.

All position sensors are limited by extraneous thermal noise. In optical-cavity-based sensors, major sources include thermomechanical and thermorefractive fluctuations of the cavity substrate⁸. These

fluctuations result in excess cavity-frequency noise, $S_{\omega}^{\text{imp,ex}}$, and limit the measurement rate to

$$\Gamma_{\text{meas}} = \frac{g_0^2/2}{S_{\omega}^{\text{imp,shot}} + S_{\omega}^{\text{imp,ex}}} = \frac{\Gamma_m/16}{n_{\text{imp}}^{\text{shot}} + n_{\text{imp}}^{\text{ex}}}$$

where $S_{\omega}^{\text{imp,shot}}$ is the photocurrent shot noise referred to an apparent cavity resonance-frequency noise, $n_{\text{imp}}^{\text{shot}} \equiv S_{\omega}^{\text{imp,shot}}/2S_{\omega}^{\text{zp}}$ and $n_{\text{imp}}^{\text{ex}} \equiv S_{\omega}^{\text{imp,ex}}/2S_{\omega}^{\text{zp}}$. Figure 1d shows the extraneous-noise floor of our sensor over a broad range of frequencies. We obtained this spectrum by subtracting shot noise from a measurement made with a large intracavity photon number, $n_c > 10^5$. (To mitigate dynamic instabilities, the measurement was conducted using approximately 10 mbar of gas pressure at an elevated temperature of 15.7 K.) High- and low- Q noise peaks correspond to thermal motion of the nanobeam and the microdisk, respectively. Near the fundamental noise peak, we observe an extraneous frequency noise of $S_{\omega}^{\text{imp,ex}} \approx (2\pi \times 30\text{ Hz}/\sqrt{\text{Hz}})^2$, corresponding to a displacement imprecision of $S_x^{\text{imp,ex}} \approx (4.3 \times 10^{-17}\text{ m}/\sqrt{\text{Hz}})^2$. We identify this imprecision as a combination of thermorefractive noise⁸, diode laser frequency noise¹⁶ and off-resonant thermal motion of the in-plane nanobeam mode at 4.6 MHz. Owing to the large zero-point motion of the oscillator, $S_{\omega}^{\text{zp}} \equiv 4g_0^2/\Gamma_m = (2\pi \times 6.7\text{ kHz}/\sqrt{\text{Hz}})^2$ (corresponding to $S_x^{\text{zp}} = (0.95 \times 10^{-14}\text{ m}/\sqrt{\text{Hz}})^2$), the thermal-phonon-equivalent magnitude of $S_x^{\text{imp,ex}}$ has an exceptionally low value of $n_{\text{imp}}^{\text{ex}} \approx 1.0 \times 10^{-5}$, nearly 44 dB below that at the SQL. The associated measurement rate, $\Gamma_m/(16n_{\text{imp}}^{\text{ex}}) \approx 2\pi \times 36\text{ kHz}$, is equal to the thermal decoherence rate at an experimentally accessible temperature of 1.3 K, which implies that ground-state cooling ($\Gamma_{\text{meas}} > \Gamma_{\text{th}}/8$) should be accessible at 10 K.

The performance of the sensor is limited by optical loss, photothermal and radiation-pressure instabilities, and extraneous sources of measurement back-action. We investigate these constraints by recording the measurement imprecision, n_{imp} , and the effective bath occupation, $n_{\text{tot}} \equiv n_{\text{th}} + n_{\text{ba}}$, as a function of intracavity photon number, and comparing their product to the uncertainty-limited value, $4\sqrt{n_{\text{imp}}n_{\text{tot}}} \geq 1$ (Fig. 2). Two considerations are crucial. First, to

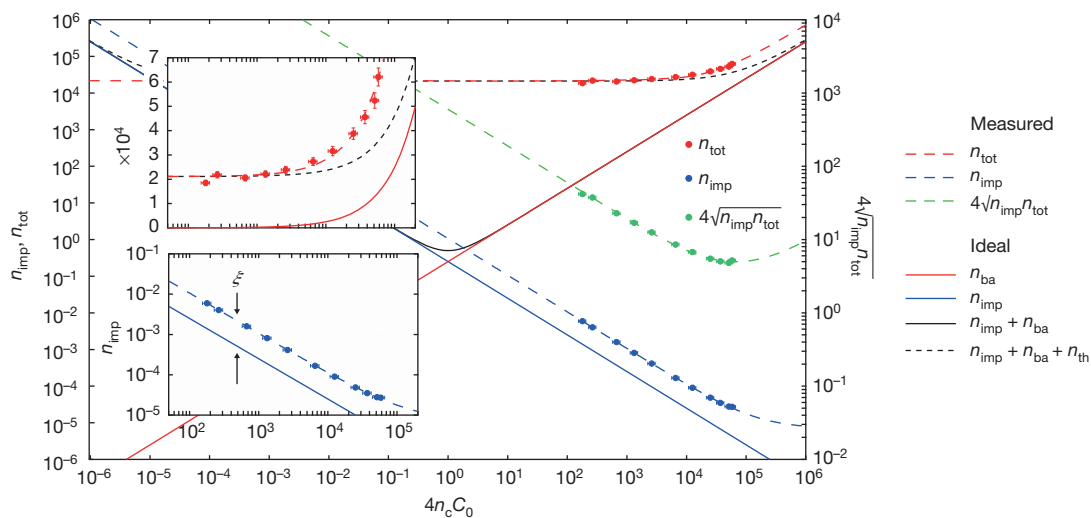


Figure 2 | Measurement imprecision and back action versus intracavity photon number. Red, blue and green points correspond to measurements of effective bath occupation, $n_{\text{tot}} = n_{\text{th}} + n_{\text{ba}}$, measurement imprecision (expressed as an equivalent thermal phonon occupation of the oscillator), n_{imp} , and effective imprecision–back-action product, $4\sqrt{n_{\text{imp}}n_{\text{tot}}}$, respectively. Solid red, blue and black lines correspond to models of ideal (lossless, quantum-limited) back-action, $n_{\text{ba}} = C_0 n_c$, imprecision, $n_{\text{imp}} = 1/(16C_0 n_c)$, and their sum, $n_{\text{ba}} + n_{\text{imp}}$, respectively. Dashed black line represents the ideal apparent

bath occupation, $n_{\text{th}} + n_{\text{ba}} + n_{\text{imp}}$. Dashed red and blue lines are models of experimental data, which include contribution from extraneous back action, $C_0^{\text{ex}} = 0.56$, extraneous imprecision, $n_{\text{imp}}^{\text{ex}} = 0.70 \times 10^{-5}$, and imperfect detection efficiency, $\xi = 0.23$, as described in the main text. Dashed green line is a model of the effective imprecision–back-action product using equation (2). Insets highlight the measurement region and have the same axes labels as the main plot. Vertical and horizontal error bars are due to systematic uncertainties in n_{th} and $C_0 n_c$, respectively (see Supplementary Information).

efficiently collect photons from the cavity, it is necessary to increase the taper-cavity coupling rate to $\kappa_{\text{ex}} \gtrsim \kappa_0$, thereby increasing the cavity decay rate to $\kappa = \kappa_0 + \kappa_{\text{ex}}$. We operate at a near-critically coupled ($\kappa_{\text{ex}} = \kappa_0$) value of $\kappa \approx 2\pi \times 0.91$ GHz, which reduces the single photon cooperativity to $C_0 \approx 0.31$ in exchange for an output coupling efficiency of $\eta_c = (\kappa - \kappa_0)/\kappa \approx 0.52$. Second, to minimize S_x^{imp} , it is necessary to maximize n_c while avoiding dynamic instabilities. We accomplish this by damping the oscillator using radiation pressure feedback. Feedback is performed by modulating the intensity of the feedback laser with an electronically amplified and delayed copy of the homodyne photocurrent. A feedback phase of $-3\pi/2$ is chosen by tuning the total delay, $\tau \approx 172$ ns, such that the feedback-induced spring effect is minimized²⁷ (see Supplementary Information). The resulting viscous radiation pressure gives rise to a well-known cooling effect ('cold-damping'^{10,17}), which reduces the phonon occupancy of the mechanical mode to a mean value of $n_m \approx n_{\text{tot}}\Gamma_m/(\Gamma_m + \Gamma_{\text{fb}})$, where Γ_{fb} is the optically induced damping rate.

Measurements of the thermal motion of the oscillator are shown in Fig. 1c. We determine n_{tot} and n_{imp} by fitting each noise peak to a Lorentzian with a linewidth of $\Gamma_{\text{eff}} = \Gamma_m + \Gamma_{\text{fb}} + \Gamma_{\text{ba}}$ (which includes a minor contribution from dynamic back action, Γ_{ba} ; see Supplementary Information), a peak amplitude of $S_\omega(\Omega_m) \approx 2n_{\text{tot}}(\Gamma_m/\Gamma_{\text{eff}})^2 S_\omega^{\text{zp}}$ and an offset of $S_\omega^{\text{imp}} = 2n_{\text{imp}}S_\omega^{\text{zp}}$. For low input powers, $n_c \ll n_{\text{th}}/C_0$, we observe that the effective bath occupation is dominated by thermalization to the cryostat, $n_{\text{tot}} \approx n_{\text{th}}$, and that imprecision scales as $n_{\text{imp}} = (16\xi C_0 n_c)^{-1}$ where $\xi \approx 0.23$. ξ represents the ideality of the measurement, and includes optical losses as well as reduction of optomechanical transduction due to cavity mode splitting (see Supplementary Information). When operating with the higher input power—ultimately limited by dynamic instability of higher-order beam modes—the lowest imprecision we observe is $n_{\text{imp}} = 2.7(\pm 0.2) \times 10^{-5}$, which corresponds to an imprecision 39.7 ± 0.3 dB below that at the SQL (see Supplementary Information for the budget of uncertainties). The associated measurement rate, $\Gamma_{\text{meas}} = 2\pi \times (13 \pm 1)$ kHz, is a factor of 9.2 lower than the rate of thermal decoherence to the ambient 4.4 K bath, $\Gamma_{\text{th}} = 2\pi \times 120$ kHz. Notably, this value is within 15% of the requirement for ground-state cooling.

For large measurement strengths, quantum back-action²⁵ should in principle exceed the ambient thermal force, scaling as $n_{\text{ba}} = C_0 n_c$ (see Supplementary Information). As shown in Fig. 2 (red data), our system deviates from this ideal behaviour, owing to excess back-action. The back-action manifests as an extraneous cooperativity, $C_0^{\text{ex}} \approx 0.56$, and limits the fractional contribution of quantum back-action to $C_0/(C_0 + C_0^{\text{ex}}) \approx 35\%$. Similar behaviour for high-order mechanical modes suggests that photothermal heating is the cause of this excess heating, as does our observation that C_0^{ex} is markedly higher at lower cryostat temperatures, which is consistent with the reduction of thermal conductivity in amorphous glasses below 10 K (see Supplementary Information). Including extraneous back-action, we model the effective imprecision–back-action product (green curve in Fig. 2) as

$$4\sqrt{n_{\text{imp}}n_{\text{tot}}} = \sqrt{\frac{1}{\xi} \left(1 + \frac{n_c}{n_c^{\text{ex}}}\right) \left(1 + \frac{n_{\text{th}}}{C_0 n_c} + \frac{C_0^{\text{ex}}}{C_0}\right)} \quad (2)$$

where $n_c^{\text{ex}} \equiv (16\xi C_0 n_{\text{imp}}^{\text{ex}})^{-1}$ is the photon number for which $n_{\text{imp}}^{\text{shot}} = n_{\text{imp}}^{\text{ex}}$. Using $n_c \approx 5 \times 10^4 \ll n_c^{\text{ex}}$, we observe a minimum value of $4\sqrt{n_{\text{imp}}n_{\text{tot}}} \approx 5.0$, which corresponds to a measurement efficiency of $\Gamma_{\text{meas}}/\Gamma_{\text{tot}} \approx 0.040$.

To illustrate the utility of this high efficiency, we increase the strength of the feedback used to cold-damp the oscillator in Fig. 1c. The limits of cold-damping are well-studied^{12,17,20}. Ignoring back-action from the weakly driven ($n_c < 100$) feedback optical mode, the phonon occupancy of the cooled mechanical mode depends on the balance between its coupling to thermal, measurement and feedback reservoirs at rates Γ_{th} , $\Gamma_m n_{\text{ba}}$ and $g_{\text{fb}}\Gamma_m n_{\text{imp}}$, respectively, where $g_{\text{fb}} \equiv \Gamma_{\text{fb}}/\Gamma_m$ is the open loop feedback gain (see Supplementary Information):

$$n_m + \frac{1}{2} = \frac{1}{1 + g_{\text{fb}}} n_{\text{tot}} + \frac{g_{\text{fb}}^2}{1 + g_{\text{fb}}} n_{\text{imp}} \geq 2\sqrt{n_{\text{imp}}n_{\text{tot}}} \quad (3)$$

(here we assume $\Gamma_{\text{ba}} \ll \Gamma_{\text{fb}}$ and $n_{\text{tot}} \gg 1$). The photon occupancy n_m is minimized for an optimal gain of $g_{\text{fb}} = \sqrt{n_{\text{tot}}/n_{\text{imp}}}$, which corresponds to suppressing S_x to S_x^{imp} (black curve in the inset of Fig. 3).

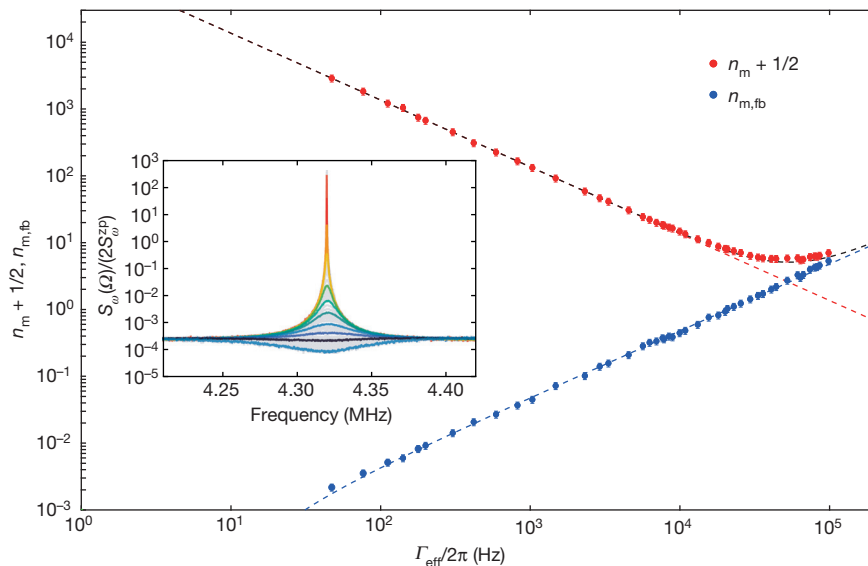


Figure 3 | Radiation-pressure feedback cooling to near the ground state. Red and blue points correspond to measurements of the phonon occupancy of the mechanical mode, n_m (plus a phonon-equivalent zero-point energy of $1/2$), and its component due to feedback of measurement noise, $n_{m,\text{fb}} = n_{\text{imp}}g_{\text{fb}}^2/(1 + g_{\text{fb}})$, respectively, versus measured damping rate, $\Gamma_{\text{eff}} = (1 + g_{\text{fb}})\Gamma_m$. Red, blue and black dashed lines correspond to models

of the components in equation (3): $n_{\text{tot}}/(1 + g_{\text{fb}})$, $n_{m,\text{fb}}$, and $n_m + 1/2$, respectively, using experimental parameters $\Gamma_m = 2\pi \times 5.7$ Hz, $n_{\text{tot}} = 2.4 \times 10^5$ and $n_{\text{imp}} = 2.9 \times 10^{-4}$. Inset, thermomechanical noise spectra for various feedback gain settings (represented by the differently coloured curves); Lorentzian fits to these spectra were used to infer the red and blue points. Error bars are due to systematic uncertainty in n_{tot} (see Supplementary Information).

Notably, for $C_0^{\text{ex}} = 0$, $n_m < 1$ requires $n_{\text{imp}} < 1/(2n_{\text{th}})$. The results presented in Fig. 2 suggest that $n_m \approx 2$ should be achievable.

Figure 3 shows the results of feedback cooling using a measurement with an imprecision far below that at the SQL. We emphasize that for this demonstration, imprecision was deliberately limited to $n_{\text{imp}} \approx 2.9 \times 10^{-4}$ to reduce uncertainties due to extraneous heating and the off-resonant tail of the thermal noise peak at 4.6 MHz. (The latter restriction limits the applicability of equation (3) to effective damping rates of $\Gamma_{\text{eff}} \approx (1 + g_{\text{fb}})\Gamma_m \lesssim 2\pi \times 200$ kHz.) Feedback gain was controlled by changing the magnitude of the electronic gain, leaving all other parameters (for example, laser power) fixed. The Markovianity condition $\tau \ll 2\pi/\Gamma_{\text{th}}$ is satisfied by the feedback delay¹². By fitting the closed-loop noise spectrum (Fig. 3, inset) to a standard Lorentzian noise-squashing model²⁸ (see Supplementary Information), we estimate the phonon occupancy of the mechanical mode using $n_m + 0.5 \approx \Gamma_{\text{eff}}(S_o(\Omega_m) + S_o^{\text{imp}})/(2S_o^{\text{zp}})$, where S_o^{imp} denotes the off-resonant background. After accounting for extraneous back-action, we infer a minimum occupation of $n_m \approx 5.3 \pm 0.6$ (see Supplementary Information for budget of uncertainties) at an optimal damping rate of $\Gamma_{\text{eff}} \approx 2\pi \times 52$ kHz, which corresponds to a ground-state probability of $1/(1 + n_m) \approx 16\%$. This value agrees well with the prediction based on equation (3) and the data shown in Fig. 2. Notably, for larger feedback strengths, shot-noise ‘squashing’²² leads to an apparent reduction of S_o , even though n_m physically increases. This discrepancy can be resolved with an out-of-loop measurement¹², at the cost of reduced measurement efficiency.

Collectively, our results establish new benchmarks for the linear measurement and control of a mechanical oscillator. Using an optomechanical sensor with a readout imprecision that is nearly 40 dB below that at the SQL, we have shown that traditional radiation pressure cold-damping¹⁰ can be used to cool a nanomechanical oscillator to a mean phonon occupancy of approximately 5.3; this represents a 40-fold improvement over previous reports^{11,28,29}, and invites comparison^{20,30} to the recent success of autonomous feedback (sideband) cooling^{21–23}. Looking forward, high-efficiency optomechanical sensors may enable a variety of feedback applications such as back-action evasion²⁴ and mechanical squeezing¹⁸.

Received 17 October 2014; accepted 3 June 2015.

Published online 10 August 2015.

1. Wiseman, H. M. Quantum theory of continuous feedback. *Phys. Rev. A* **49**, 2133–2150 (1994); erratum **49**, 5159 (1994).
2. Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2009).
3. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477**, 73–77 (2011).
4. Vijay, R. *et al.* Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature* **490**, 77–80 (2012).
5. Anetsberger, G. *et al.* Near-field cavity optomechanics with nanomechanical oscillators. *Nature Phys.* **5**, 909–914 (2009).
6. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82**, 1155–1208 (2010).
7. Teufel, J. D., Donner, T., Castellanos-Beltran, M. A., Harlow, J. W. & Lehnert, K. W. Nanomechanical motion measured with an imprecision below that at the standard quantum limit. *Nature Nanotechnol.* **4**, 820–823 (2009).
8. Anetsberger, G. *et al.* Measuring nanomechanical motion with an imprecision below the standard quantum limit. *Phys. Rev. A* **82**, 061804(R) (2010).
9. Westphal, T. *et al.* Interferometer readout noise below the standard quantum limit of a membrane. *Phys. Rev. A* **85**, 063806 (2012).
10. Cohadon, P. F., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).

11. LIGO Scientific Collaboration. Observation of a kilogram-scale oscillator near its quantum ground state. *New J. Phys.* **11**, 073032 (2009).
12. Bushev, P. *et al.* Feedback cooling of a single trapped ion. *Phys. Rev. Lett.* **96**, 043003 (2006).
13. D’Urso, B., Odom, B. & Gabrielse, G. Feedback cooling of a one-electron oscillator. *Phys. Rev. Lett.* **90**, 043001 (2003).
14. Hatridge, M. *et al.* Quantum back-action of an individual variable-strength measurement. *Science* **339**, 178–181 (2013).
15. Caves, C. M. Quantum-mechanical radiation-pressure fluctuations in an interferometer. *Phys. Rev. Lett.* **45**, 75–79 (1980).
16. Aspelmeier, M., Kippenberg, T. J. & Marquardt, F. Cavity optomechanics. *Rev. Mod. Phys.* **86**, 1391–1452 (2014).
17. Courty, J.-M., Heidmann, A. & Pinard, M. Quantum limits of cold damping with optomechanical coupling. *Eur. Phys. J. D* **17**, 399–408 (2001).
18. Szorkovszky, A., Doherty, A. C., Harris, G. I. & Bowen, W. P. Mechanical squeezing via parametric amplification and weak measurement. *Phys. Rev. Lett.* **107**, 213603 (2011).
19. Mancini, S., Vitali, D. & Tombesi, P. Optomechanical cooling of a macroscopic oscillator by homodyne feedback. *Phys. Rev. Lett.* **80**, 688–691 (1998).
20. Genes, C., Vitali, D., Tombesi, P., Gigan, S. & Aspelmeier, M. Ground-state cooling of a micromechanical oscillator: comparing cold damping and cavity-assisted cooling schemes. *Phys. Rev. A* **77**, 033804 (2008); erratum **79**, 039903 (2009).
21. Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
22. Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
23. Verhagen, E., Deléglise, S., Weis, S., Schliesser, A. & Kippenberg, T. J. Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode. *Nature* **482**, 63–67 (2012).
24. Wiseman, H. M. Using feedback to eliminate back-action in quantum measurements. *Phys. Rev. A* **51**, 2459–2468 (1995).
25. Purdy, T. P., Peterson, R. W. & Regal, C. A. Observation of radiation pressure shot noise on a macroscopic object. *Science* **339**, 801–804 (2013).
26. Murch, K. W., Moore, K. L., Gupta, S. & Stamper-Kurn, D. M. Observation of quantum-measurement backaction with an ultracold atomic gas. *Nature Phys.* **4**, 561–564 (2008).
27. Gavartin, E., Verlot, P. & Kippenberg, T. J. A hybrid on-chip optomechanical transducer for ultrasensitive force measurements. *Nature Nanotechnol.* **7**, 509–514 (2012).
28. Poggio, M., Degen, C., Mamin, H. & Rugar, D. Feedback cooling of a cantilever’s fundamental mode below 5 mK. *Phys. Rev. Lett.* **99**, 017201 (2007).
29. Li, T., Kheifets, S. & Raizen, M. G. Millikelvin cooling of an optically trapped microsphere in vacuum. *Nature Phys.* **7**, 527–530 (2011).
30. Jacobs, K., Nurdin, H. I., Strauch, F. W. & James, M. Comparing resolved-sideband cooling and measurement-based feedback cooling on an equal footing: analytical results in the regime of ground-state cooling. *Phys. Rev. A* **91**, 043812 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge nanofabrication advice from E. Gavartin in the early stages of the project. All samples were fabricated at the CMI (Center for Micro-Nanotechnology) at EPFL. Research was funded by an ERC Advanced Grant (QuREM), by the DARPA/MTO ORCHID programme, the Marie Curie Initial Training Network ‘Cavity Quantum Optomechanics’ (cQOM), the Swiss National Science Foundation and through support from the NCCR of Quantum Engineering (QSIT). N.P. and D.J.W. acknowledge support from the European Commission through Marie Skłodowska-Curie Fellowships: IEF project 303029 and IIF project 331985, respectively.

Author Contributions R.S., A.G. and D.J.W. contributed to the design and initial characterization of the device. R.S. developed, implemented and optimized the fabrication process. A.G. developed and implemented a numerical model to aid in device optimization. D.J.W., V.S. and N.P. conceived of, designed and performed the experiment. V.S. and D.J.W. analysed the data with support from N.P. V.S. developed the theoretical framework of the experiment with support from D.J.W. and N.P. V.S. wrote sections I and II of the Supplementary Information. R.S., V.S., D.J.W. and A.G. wrote section III of the Supplementary Information. D.J.W., N.P. and V.S. wrote sections IV and V of the Supplementary Information. D.J.W. wrote the main text with support from all other authors. T.J.K. oversaw all aspects of the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.J.K. (tobias.kippenberg@epfl.ch).

Switching on elusive organometallic mechanisms with photoredox catalysis

Jack A. Terrett¹, James D. Cuthbertson¹, Valerie W. Shurtleff¹ & David W. C. MacMillan¹

Transition-metal-catalysed cross-coupling reactions have become one of the most used carbon–carbon and carbon–heteroatom bond-forming reactions in chemical synthesis. Recently, nickel catalysis has been shown to participate in a wide variety of C–C bond-forming reactions, most notably Negishi, Suzuki–Miyaura, Stille, Kumada and Hiyama couplings^{1,2}. Despite the tremendous advances in C–C fragment couplings, the ability to forge C–O bonds in a general fashion via nickel catalysis has been largely unsuccessful. The challenge for nickel-mediated alcohol couplings has been the mechanistic requirement for the critical C–O bond-forming step (formally known as the reductive elimination step) to occur via a Ni(III) alkoxide intermediate. Here we demonstrate that visible-light-excited photoredox catalysts can modulate the preferred oxidation states of nickel alkoxides in an operative catalytic cycle, thereby providing transient access to Ni(III) species that readily participate in reductive elimination. Using this synergistic merger of photoredox and nickel catalysis, we have developed a highly efficient and general carbon–oxygen coupling reaction using abundant alcohols and aryl bromides. More notably, we have developed a general strategy to ‘switch on’ important yet elusive organometallic mechanisms via oxidation state modulations using only weak light and single-electron-transfer catalysts.

Visible-light-mediated photoredox catalysis has gained momentum over the past decade as a platform for the development of novel synthetic transformations via the implementation of non-traditional open-shell mechanisms. This catalysis field employs transition metal polypyridyl complexes or organic dyes that, upon excitation by visible light, readily engage in an array of single-electron transfer (SET) processes^{3–5} that often include oxidation, reduction or redox pathways that have previously been elusive^{6–9}. Indeed, several research groups have demonstrated that photoredox catalysis can be combined with transition metal catalysis to achieve a series of unique bond-forming reactions that take advantage of the known reactivity of each individual mode of catalysis^{10–16}. We recently questioned whether the combination of photoredox catalysis and transition metal catalysis might be capable of delivering fundamentally new organometallic reactivity by providing access to currently unknown or inaccessible mechanistic pathways.

In a series of seminal studies, Hillhouse and colleagues demonstrated that Ni(II) alkoxide complexes do not readily undergo reductive elimination at ambient or elevated temperature and that stoichiometric conversion to a less stable Ni(III) system is required for productive C–O bond formation^{17–19}. Similarly, C–heteroatom reductive elimination has been demonstrated when stoichiometrically accessing Ni(IV) intermediates^{20,21}. In fact, computational studies have revealed that Ni(II) alkoxide reductive elimination is endothermic, in contrast to Pd, Pt or alkyl Ni(II) variants, which are exothermic (Fig. 1)²². Indeed, the use of palladium and copper catalysis to generate aryl ethers from alcohols is well preceded and broadly employed^{23–25}. The utilization of nickel catalysis in C–O bond construction would be an important complementary method, considering the diversity of electrophiles amenable to nickel cross-couplings, such as alkyl halides

and aryl pseudohalides^{1,2}. Hartwig and colleagues have reported a Ni(COD)₂-catalysed aryl etherification at elevated temperature using three preformed oxides, namely sodium methoxide, ethoxide and *tert*-butyldimethylsiloxide with electron-deficient aryl bromides²⁶. However, a general nickel-catalysed C–O coupling strategy under mild conditions has thus far been elusive and no examples with alcohols have been documented. We recently questioned whether photoredox catalysis might be used to modulate the range of nickel alkoxide oxidation states that can be accessed during a conventional catalytic cycle. More specifically, we recognized that it should be possible to utilize the photonic energy of weak visible light to expand thermodynamically the number of nickel-catalyst oxidation states via two discrete photocatalyst SET events: first, oxidation of Ni(II) to the elusive Ni(III)–alkoxide complex (which we assumed would facilitate the critical C–O bond-forming step); and second, reduction of Ni(III) to Ni(II), to enable the subsequent aryl bromide oxidative addition steps. If successful, we recognized that this new synergistic catalysis pathway might provide (1) the first general example of a nickel-catalysed C–O coupling reaction employing simple alcohols and aryl halides, and

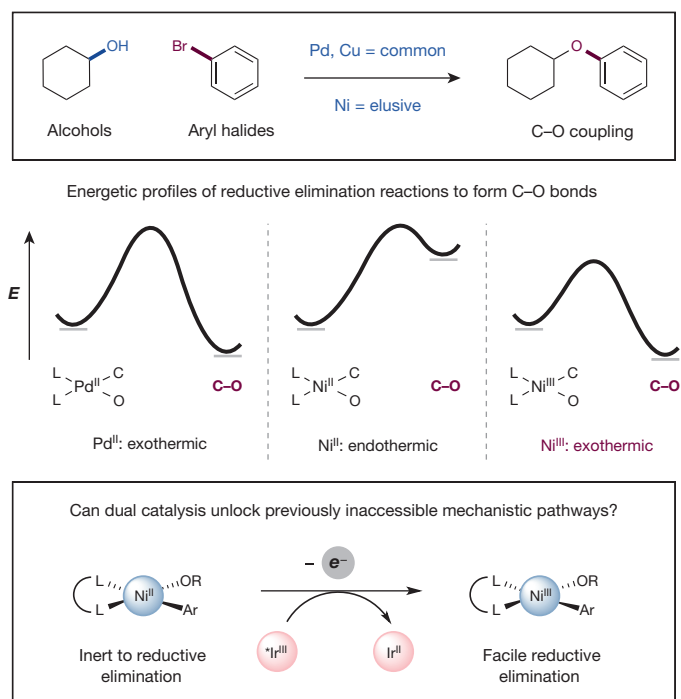


Figure 1 | Modulating oxidation states of nickel enables challenging carbon–heteroatom coupling. Pd(II)-catalysed C–O reductive elimination is an exothermic process and well preceded. Ni(II) C–O reductive elimination is thermodynamically disfavoured—we postulated that accessing Ni(III) by photoredox-mediated oxidation state manipulation could switch on C–O coupling in a general fashion, circumventing this thermodynamic restriction.

¹Merck Center for Catalysis at Princeton University, Princeton, New Jersey 08544, USA.

(2) a demonstration of a new strategy by which elusive organometallic couplings can be ‘switched on’ via oxidation state modulation.

A detailed description of our proposed mechanistic cycle for the photoredox/nickel-catalysed C–O coupling is outlined in Fig. 2a. Oxidative addition of Ni(0) (species 1) into an aryl bromide would deliver a Ni(II) aryl complex 2 (ref. 27). At this stage, ligand exchange (displacement of the bromide ion with the substrate alcohol) would produce the Ni(II) aryl alkoxide (3)—an organometallic species that traditionally represents a catalytic ‘dead end’. At the same time, visible light irradiation of heteroleptic iridium(III) photocatalyst Ir[dF(CF₃)ppy]₂(dtbbpy)PF₆ [dF(CF₃)ppy = 2-(2,4-difluorophenyl)-5-(trifluoromethyl)pyridine, dtbbpy = 4,4′-di-*tert*-butyl-2,2′-bipyridine] (4) would produce the long-lived photoexcited *Ir^{III} state 5 (excited-state lifetime τ = 2.3 μ s) (ref. 28). At this juncture, we hypothesized that the nickel and photoredox cycles would merge via SET between the

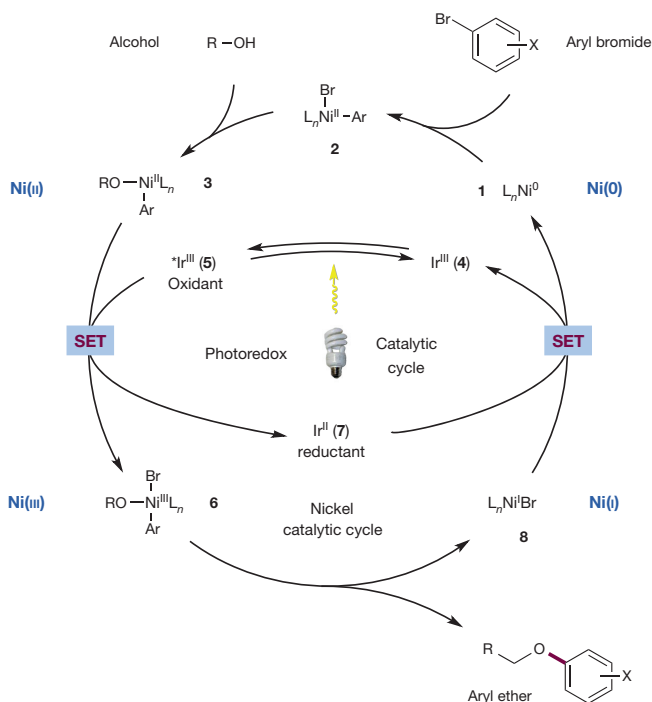
Ni(II) complex 3 and the highly oxidizing photoexcited *Ir(III) 5 to generate the critical Ni(III) aryl alkoxide 6 and the reduced Ir(II) photocatalyst 7. On the basis of the established redox potentials of these catalytic intermediates, we envisioned this key electron transfer step to be kinetically and thermodynamically favourable under standard reaction conditions (half-reaction reduction potential $E_{1/2}^{\text{red}}$ [*Ir^{III}/Ir^{II}] = +1.21 V versus the saturated calomel electrode (SCE) in CH₃CN, $E_{1/2}^{\text{red}}$ [Ni^{III}/Ni^{II}] = +0.71 V versus Ag/AgCl in CH₂Cl₂ for (bpy)Ni^{II}(Mes)OMe)^{28,29}. Once formed, we assumed the transient Ni(III) complex 6 would rapidly undergo reductive elimination to forge the desired C–O bond, while delivering the aryl ether product and Ni(I) complex 8. At this stage, we envisioned both catalytic cycles merging for a second time, enabling the single-electron reduction of Ni(I) to Ni(0) by Ir(II) species 7 ($E_{1/2}^{\text{red}}$ [Ir^{III}/Ir^{II}] = –1.37 V versus SCE in CH₃CN), thereby completing the nickel and photoredox catalytic cycles at the same moment²⁸.

On the basis of this synergistic catalysis design plan, we began our investigations into the proposed nickel-catalysed aryl–alcohol etherification using 1-hexanol and 4-bromoacetophenone. As expected, the use of traditional nickel cross-coupling conditions (that is, without photoredox), led to no observable product despite the implementation of a range of Ni(II) and Ni(0) complexes (Fig. 2b, equation (1)). By contrast, we found that the desired C–O bond could be forged in excellent yield (86%) at room temperature via the introduction of the photocatalyst Ir[dF(CF₃)ppy]₂(dtbbpy)PF₆ in the presence of Ni(COD)₂, dtbbpy, quinuclidine, K₂CO₃ and blue light-emitting diodes (LEDs) as the visible light source. Given the inherent advantages of employing a bench-stable Ni(II) catalyst in lieu of Ni(COD)₂, we hypothesized that a catalytically active Ni(0) species 1 might be accessible *in situ* via two SET reductions of (dtbbpy)Ni(II)Cl₂ using the iridium photocatalyst ($E_{1/2}^{\text{red}}$ [Ir^{III}/Ir^{II}] = –1.37 V versus SCE in CH₃CN, $E_{1/2}^{\text{red}}$ [Ni^{II}/Ni⁰] = –1.2 V versus SCE in *N,N*-dimethylformamide (DMF))^{28,30}. In this vein, we hypothesized that quinuclidine may also serve as a sacrificial reductant and were pleased to observe that with NiCl₂(dtbbpy) the desired fragment etherification was accomplished in 91% yield (Fig. 2b, equation (2)).

With the optimal conditions in hand, we next sought to explore the scope of the aryl bromide component in this new nickel-catalysed C–O coupling reaction. As shown in Fig. 3, a diverse array of electron-deficient bromoarenes with a variety of functional groups (ketones, trifluoromethyls, nitriles, sulfones, esters) perform well using this synergistic protocol (compounds 9–14, 88–96% yield). Moreover, 3-bromoanisole was found to be a competent substrate in this transformation, demonstrating the diversity of substituents that can be tolerated (26, 80% yield). Notably, bicyclic aromatics such as phthalimides and phthalides couple with high levels of efficiency (15 and 16, 92% and 80% yield, respectively). With respect to heteroaromatic coupling partners, we have found that a range of pyridines with substitution at the 2 and 3 positions (nitrile, trifluoromethyl, methyl) are effective electrophiles in this protocol (17–20, 67–91% yield). Furthermore, quinolines, azaindoles and pyrimidines can be employed without loss in efficiency (21–23, 60–88% yield). Perhaps most importantly, this transformation is not limited to electron-deficient aryl bromides. For example, phenyl and *tert*-butylphenyl can be readily incorporated in this nickel etherification (24 and 25, 68% and 77% yield, respectively). It should be noted that comparable yields of aryl ether product 13 were achieved when either coupling partner was employed in excess (1.5 equiv.) (see Supplementary Information).

We next turned our attention to the alcohol reaction component. As shown in Fig. 3, we discovered that a host of primary alcohols were effective coupling partners, including substrates that incorporate benzyl, *tert*-butyl, cyclopropyl and alkene functionalities (27–30, 77–89% yield). Moreover, carbamates and ester groups are tolerated, as exemplified by the etherification of a serine derivative in 78% yield (31). Interestingly, trifluoroethanol couples proficiently in this reaction despite its diminished nucleophilicity (32, 77% yield). Both

a Modulating nickel oxidation states with photoredox catalysis



b Ni catalysis alone; equation (1)

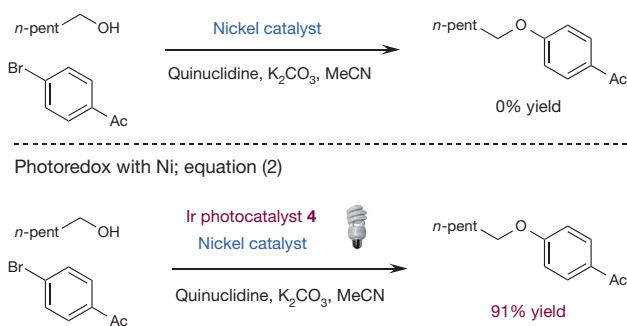


Figure 2 | Proposed mechanism by which photoredox catalysis switches on challenging nickel-catalysed C–O coupling. **a**, The catalytic cycle begins with oxidative addition of Ni(0) (species 1) into an aryl bromide to give Ni(II) aryl 2. Ligand exchange with an alcohol under basic conditions produces Ni(II) aryl alkoxide 3. Excitation of photocatalyst 4 gives the excited state 5, which can then oxidize 3 to the key Ni(III) intermediate 6 and generate Ir(II) 7. Ni(III) 6 readily reductively eliminates to form the aryl ether product and Ni(I) 8. A second SET event closes the two catalytic cycles, regenerating Ni(0) 1 and Ir(III) 4. **b**, Nickel-catalysed C–O reductive elimination can be turned on by addition of a photoredox catalyst and visible light.

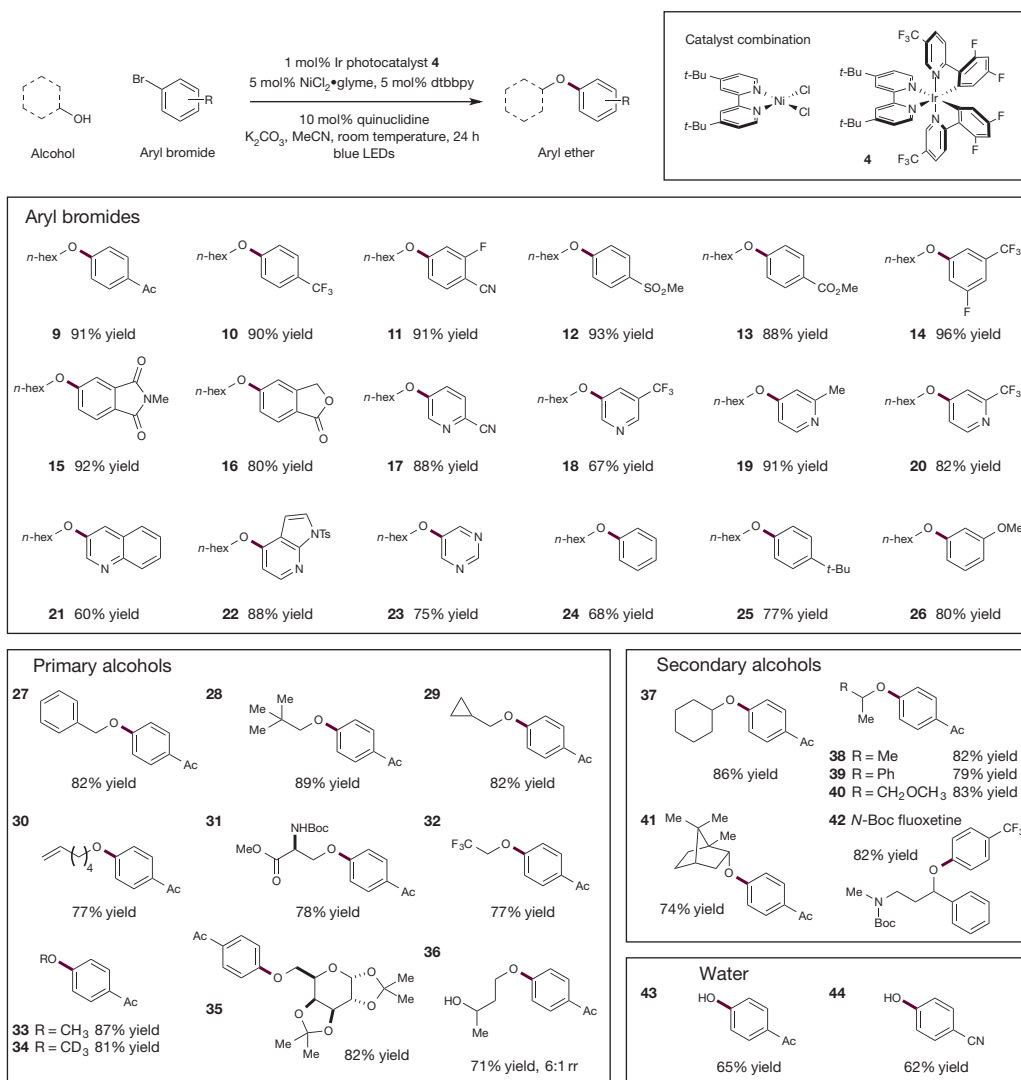


Figure 3 | Alcohol and aryl halide scope in the photoredox-nickel-catalysed C–O coupling reaction. A broad range of aryl bromides and alcohols are efficiently coupled to produce aryl ethers under the standard reaction conditions (top, generalized reaction). The aryl bromide scope includes electron-deficient and electron-neutral arenes and heteroarenes with diverse

functionalities. Both primary and secondary alcohols are proficient coupling partners under the standard conditions. Water can be employed as the nucleophile to generate phenol derivatives in a single step. Isolated yields are indicated below each entry. See Supplementary Information for experimental details.

methanol and *d*₄-methanol are effective substrates, providing a valuable strategy for installing H₃CO– and D₃CO– groups on arenes (**33** and **34**, 87% and 81% yield, respectively).

Relatively complex substrates are also tolerated in this transformation, as demonstrated by the etherification of a protected pyranose in excellent yield (**35**, 82% yield). Notably, secondary alcohols are equally effective in this etherification protocol—alkyl, benzyl and protected ether functionalities all being amenable to this technology (**37–40**, 79–86% yield). Moreover, secondary alcohols possessing β-quaternary carbons could be employed with good levels of efficiency (**41**, 74% yield). In the case of a substrate possessing both a primary and secondary alcohol, the less hindered site is arylated predominantly, as demonstrated by product **36** (71% yield, 6:1 regiomer ratio (rr)). This nickel-catalysed process can be applied to the expedient synthesis of medically relevant molecules, such as the antidepressant fluoxetine (Prozac), in three chemical steps (**42**, 82% yield for coupling step). It is important to note that using H₂O as the nucleophile delivers the corresponding phenol product in a single step from commercially available materials (**43** and **44**, 65% and 62% yield, respectively).

On the basis of our proposed design plan (Fig. 2a), we initiated mechanistic studies to determine if the critical Ni(III) oxidation state

was indeed operative in this new photoredox protocol. As shown in Fig. 4a, we preformed the stable Ni(II)(aryl)(alkoxide) complex **45** to examine whether C–O reductive elimination might be achieved under various photoredox or non-photoredox conditions. Importantly, in the absence of either light or photocatalyst, none of the desired aryl ether product **46** was observed. However, when complex **45** was exposed to Ir[dF(CF₃)ppy]₂(dtbbpy)PF₆ (**4**) and a visible light source, the desired ether product **46** was formed in 59% yield. We attribute this result to the formation of a transient Ni(III) complex **47** via the single-electron reduction of *Ir^{III}, thereby enabling the favourable Ni(III) C–O reductive elimination, a result that is consistent with the seminal studies of Hillhouse. Furthermore, cyclic voltammetry of complex **45** showed an irreversible oxidation at +0.83 V versus SCE in CH₃CN (Fig. 4b), which we attribute to the oxidation of Ni(II) to Ni(III) (**45** to **47**). As such, we expect the oxidation of Ni(II) aryl alkoxides such as **3** to be readily accomplished by *Ir^{III} photocatalyst **5** (*E*_{1/2}^{red} [*Ir^{III}/Ir^{II}] = +1.21 V versus SCE)²⁸. Finally, Stern–Volmer fluorescence quenching experiments have demonstrated that the emission intensity of *Ir^{III} **5** is diminished in the presence of Ni(II) complex **45**, presumably signifying an oxidation event to form Ni(III) **47** (see Supplementary Fig. 7).

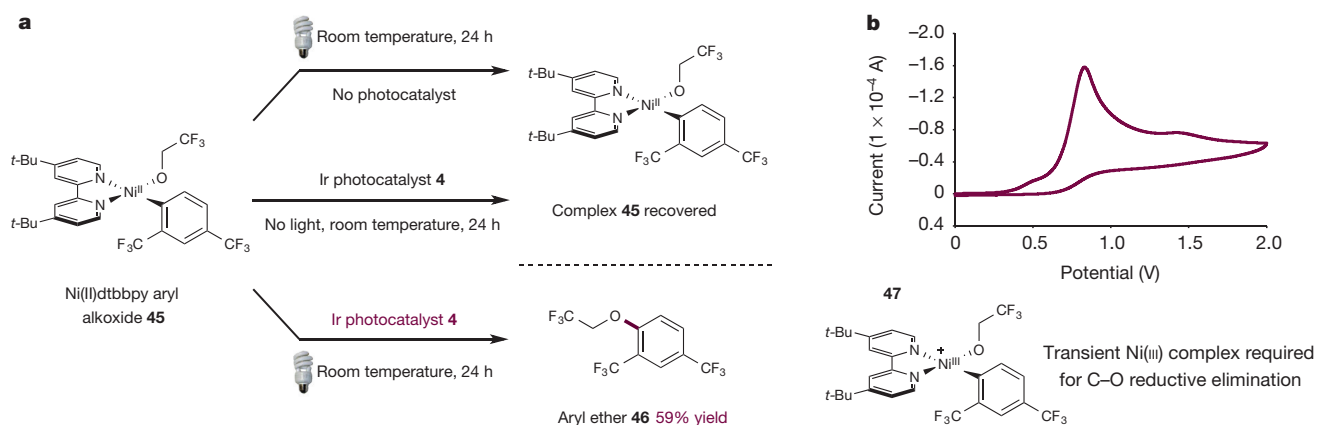


Figure 4 | Mechanistic studies support the intermediacy of transient Ni(III) complex to enable C–O reductive elimination. **a**, Reductive elimination to form C–O bond only occurs in the presence of photocatalyst and light. Reactions performed on 5.55 μ mol scale with 41 mol% photocatalyst **1** and blue

LEDs. See Supplementary Information for experimental details. **b**, Cyclic voltammogram of **45** shows Ni^{III}/Ni^{II} couple at +0.83 V versus SCE in CH₃CN with 0.1 M tetrabutylammonium hexafluorophosphate as the supporting electrolyte at 100 mV s^{−1}.

A further series of experiments were performed using Ni(COD)₂ as the nickel catalyst in lieu of NiCl₂ (see Supplementary Fig. 1). In the presence of this Ni(0) catalyst, the reaction proceeds with yields that are comparable to the Ni(II) precatalyst; however, when the iridium photocatalyst is omitted, the desired C–O coupling is not observed. This result suggests again that reductive elimination does not occur from the Ni(II) oxidation state as a single-electron oxidation to Ni(III) is required to enable product formation. Interestingly, when quinuclidine is omitted, the reaction efficiency is greatly diminished (8% yield over 24 h). This effect can possibly be attributed to the amine functioning as an electron shuttle, facilitating reduction of Ni(I) and oxidation of Ni(II); however, mechanistic studies are currently ongoing to elucidate the role of quinuclidine in more detail.

We have developed a catalytic strategy for accessing transient Ni(III) complexes through the use of visible-light-mediated photoredox catalysis for application in various challenging C–O cross-couplings. This method of modulating transition metal oxidation states has enabled a previously elusive transformation within the realm of nickel catalysis. We anticipate that this new mechanistic paradigm will find application in using nickel and other transition metals in a series of challenging bond constructions.

Received 6 May; accepted 22 June 2015.

Published online 12 August 2015.

- Tasker, S. Z., Standley, E. A. & Jamison, T. F. Recent advances in homogeneous nickel catalysis. *Nature* **509**, 299–309 (2014).
- Netherton, M. R. & Fu, G. C. Nickel-catalyzed cross-couplings of unactivated alkyl halides and pseudohalides with organometallic compounds. *Adv. Synth. Catal.* **346**, 1525–1532 (2004).
- Narayanan, J. M. R. & Stephenson, C. R. J. Visible light photoredox catalysis: applications in organic synthesis. *Chem. Soc. Rev.* **40**, 102–113 (2011).
- Prier, C. K., Rankic, D. A. & MacMillan, D. W. C. Visible light photoredox catalysis with transition metal complexes: applications in organic synthesis. *Chem. Rev.* **113**, 5322–5363 (2013).
- Schultz, D. M. & Yoon, T. P. Solar synthesis: prospects in visible light photocatalysis. *Science* **343**, 1239176 (2014).
- Nicewicz, D. A. & MacMillan, D. W. C. Merging photoredox catalysis with organocatalysis: the direct asymmetric alkylation of aldehydes. *Science* **322**, 77–80 (2008).
- Ischay, M. A., Anzovino, M. E., Du, J. & Yoon, T. P. Efficient visible light photocatalysis of [2+2] enone cycloadditions. *J. Am. Chem. Soc.* **130**, 12886–12887 (2008).
- Narayanan, J. M. R., Tucker, J. W. & Stephenson, C. R. J. Electron-transfer photoredox catalysis: development of a tin-free reductive dehalogenation reaction. *J. Am. Chem. Soc.* **131**, 8756–8757 (2009).
- Pirnot, M. T., Rankic, D. A., Martin, D. B. C. & MacMillan, D. W. C. Photoredox activation for the direct β -arylation of ketones and aldehydes. *Science* **339**, 1593–1596 (2013).
- Hopkinson, M. N., Sahoo, B., Li, J.-L. & Glorius, F. Dual catalysis sees the light: combining photoredox with organo-, acid, and transition-metal catalysis. *Chemistry* **20**, 3874–3886 (2014).
- Osawa, M., Nagai, H. & Akita, M. Photo-activation of Pd-catalyzed Sonogashira coupling using a Ru/bipyridine complex as energy transfer agent. *Dalton Trans.* 827–829 (2007).
- Kalyani, D., McMurtrey, K. B., Neufeldt, S. R. & Sanford, M. S. Room-temperature C–H arylation: merger of Pd-catalyzed C–H functionalization and visible-light photocatalysis. *J. Am. Chem. Soc.* **133**, 18566–18569 (2011).
- Ye, Y. & Sanford, M. S. Merging visible-light photocatalysis and transition-metal catalysis in the copper-catalyzed trifluoromethylation of boronic acids with CF₃I. *J. Am. Chem. Soc.* **134**, 9034–9037 (2012).
- Sahoo, B., Hopkinson, M. N. & Glorius, F. Combining gold and photoredox catalysis: visible light-mediated oxy- and aminoarylation of alkenes. *J. Am. Chem. Soc.* **135**, 5505–5508 (2013).
- Tellis, J. C., Primer, D. N. & Molander, G. A. Single-electron transmetalation in organoboron cross-coupling by photoredox/nickel dual catalysis. *Science* **345**, 433–436 (2014).
- Zuo, Z. *et al.* Merging photoredox with nickel catalysis: coupling of α -carboxyl sp³-carbons with aryl halides. *Science* **345**, 437–440 (2014).
- Matsunaga, P. T., Hillhouse, G. L. & Rheingold, A. L. Oxygen-atom transfer from nitrous oxide to a nickel metallacycle. Synthesis, structure, and reactions of [cyclic] (2,2′-bipyridine)Ni(OCH₂CH₂CH₂CH₂). *J. Am. Chem. Soc.* **115**, 2075–2077 (1993).
- Matsunaga, P. T., Mavropoulos, J. C. & Hillhouse, G. L. Oxygen-atom transfer from nitrous oxide (N=N=O) to nickel alkyls. Syntheses and reactions of nickel(II) alkoxides. *Polyhedron* **14**, 175–185 (1995).
- Han, R. & Hillhouse, G. L. Carbon–oxygen reductive-elimination from nickel(II) oxametallacycles and factors that control formation of ether, aldehyde, alcohol, or ester products. *J. Am. Chem. Soc.* **119**, 8135–8136 (1997).
- Camasso, N. M. & Sanford, M. S. Design, synthesis, and carbon-heteroatom coupling reactions of organometallic nickel(IV) complexes. *Science* **347**, 1218–1220 (2013).
- Zhou, W., Schultz, J. W., Rath, N. P. & Mirica, L. M. Aromatic methoxylation and hydroxylation by organometallic high-valent nickel complexes. *J. Am. Chem. Soc.* **137**, 7604–7607 (2015).
- Macgregor, S. A., Neave, G. W. & Smith, C. Theoretical studies on C–heteroatom bond formation via reductive elimination from group 10 M(PH₃)₂(CH₃)(X) species (X = CH₃, NH₂, OH, SH) and the determination of metal–X bond strengths using density functional theory. *Faraday Discuss.* **124**, 111–127 (2003).
- Torraca, K. E., Huang, X., Parrish, C. A. & Buchwald, S. L. An efficient intermolecular palladium-catalyzed synthesis of aryl ethers. *J. Am. Chem. Soc.* **123**, 10770–10771 (2001).
- Wolter, M., Nordmann, G., Job, G. E. & Buchwald, S. L. Copper-catalyzed coupling of aryl iodides with aliphatic alcohols. *Org. Lett.* **4**, 973–976 (2002).
- Kataoka, N., Shelby, Q., Stambuli, J. P. & Hartwig, J. F. Air stable, sterically hindered ferrocenyl dialkylphosphines for palladium-catalyzed C–C, C–N, and C–O bond-forming cross-couplings. *J. Org. Chem.* **67**, 5553–5566 (2002).
- Mann, G. & Hartwig, J. F. Nickel- vs. palladium-catalyzed synthesis of protected phenols from aryl halides. *J. Org. Chem.* **62**, 5413–5418 (1997).
- Amatore, C. & Jutand, A. Rates and mechanism of biphenyl synthesis catalyzed by electrogenerated coordinatively unsaturated nickel complexes. *Organometallics* **7**, 2203–2214 (1988).
- Lowry, M. S. *et al.* Single-layer electroluminescent devices and photoinduced hydrogen production from an ionic iridium(III) complex. *Chem. Mater.* **17**, 5712–5719 (2005).

29. Klein, A. *et al.* Halide ligands—more than just σ -donors? A structural and spectroscopic study of homologous organonickel complexes. *Inorg. Chem.* **47**, 11324–11333 (2008).
30. Durandetti, M., Devaud, M. & Perichon, J. Investigation of the reductive coupling of aryl halides and/or ethylchloroacetate electrocatalyzed by the precursor $\text{NiX}_2(\text{bpy})$ with $\text{X}^- = \text{Cl}^-$, Br^- or MeSO_3^- and $\text{bpy} = 2,2'$ -dipyridyl. *New J. Chem.* **20**, 659–667 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements Financial support was provided by the National Institute of General Medical Sciences (R01 GM093213-01) and gifts from Merck, AbbVie and Bristol-Myers Squibb. J.A.T. thanks Bristol-Myers Squibb for a Graduate Fellowship.

J.D.C. thanks Marie Curie Actions for an International Outgoing Fellowship. The authors thank Eric R. Welin for assistance in preparing $\text{Ni}(\text{II})$ complexes.

Author Contributions J.A.T., J.D.C. and V.W.S. performed and analysed experiments. J.A.T., J.D.C., V.W.S. and D.W.C.M. designed experiments to develop this reaction and probe its utility, and also prepared this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.C.M. (dmacmill@princeton.edu).

Reduced carbon emission estimates from fossil fuel combustion and cement production in China

Zhu Liu^{1,2,3}, Dabo Guan^{4,5}, Wei Wei⁶, Steven J. Davis^{2,7}, Philippe Ciais⁸, Jin Bai⁹, Shushi Peng^{8,10}, Qiang Zhang⁴, Klaus Hubacek¹¹, Gregg Marland¹², Robert J. Andres¹³, Douglas Crawford-Brown¹⁴, Jintai Lin¹⁵, Hongyan Zhao⁴, Chaopeng Hong^{4,16}, Thomas A. Boden¹³, Kuishuang Feng¹¹, Glen P. Peters¹⁷, Fengming Xi^{2,18}, Junguo Liu^{19,20,21}, Yuan Li⁵, Yu Zhao²², Ning Zeng^{23,24} & Kebin He¹⁶

Nearly three-quarters of the growth in global carbon emissions from the burning of fossil fuels and cement production between 2010 and 2012 occurred in China^{1,2}. Yet estimates of Chinese emissions remain subject to large uncertainty; inventories of China's total fossil fuel carbon emissions in 2008 differ by 0.3 gigatonnes of carbon, or 15 per cent^{1,3–5}. The primary sources of this uncertainty are conflicting estimates of energy consumption and emission factors, the latter being uncertain because of very few actual measurements representative of the mix of Chinese fuels. Here we re-evaluate China's carbon emissions using updated and harmonized energy consumption and clinker production data and two new and comprehensive sets of measured emission factors for Chinese coal. We find that total energy consumption in China was 10 per cent higher in 2000–2012 than the value reported by China's national statistics⁶, that emission factors for Chinese coal are on average 40 per cent lower than the default values recommended by the Intergovernmental Panel on Climate Change⁷, and that emissions from China's cement production are 45 per cent less than recent estimates^{1,4}. Altogether, our revised estimate of China's CO₂ emissions from fossil fuel combustion and cement production is 2.49 gigatonnes of carbon (2 standard deviations = ± 7.3 per cent) in 2013, which is 14 per cent lower than the emissions reported by other prominent inventories^{1,4,8}. Over the full period 2000 to 2013, our revised estimates are 2.9 gigatonnes of carbon less than previous estimates of China's cumulative carbon emissions^{1,4}. Our findings suggest that overestimation of China's emissions in 2000–2013 may be larger than China's estimated total forest sink in 1990–2007 (2.66 gigatonnes of carbon)⁹ or China's land carbon sink in 2000–2009 (2.6 gigatonnes of carbon)¹⁰.

Reports of national carbon emissions^{7,11–14} are based on activity data (that is, amounts of fuels burned) and emission factors (that is, amount of carbon oxidized per unit of fuel consumed), with these factors estimated as the product of the net carbon content (that is, tonnes carbon per joule), net heating value (that is, joules per tonne fuel), total carbon content (that is, tonnes carbon per tonne fuel) and oxidation rate (that is, carbon oxidized per carbon content; see Methods). The uncertainty of China's emissions estimates is typically reported as $\pm 5\%$ to $\pm 10\%$

(refs 4, 13, 15), but this range is somewhat arbitrary because neither the activity data nor the accuracy of emission factors is well known. For instance, national activity data are substantially different from the sum of provincial activity data¹⁶, and the emissions factors used are not based on up-to-date measurements of the fuels actually being burned in China, of which the quality and mix are known to vary widely from

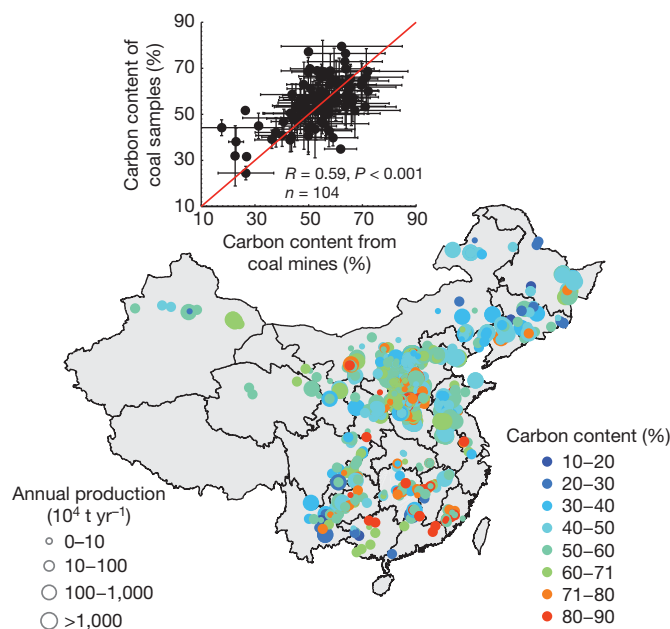


Figure 1 | Total carbon content and production of coal mines. The inset shows the comparison between carbon content from 602 coal samples and 4,243 coal mines ($R = 0.59$, $P < 0.001$, $n = 104$). Each dot in the inset indicates the average of carbon content from 602 coal samples and 4,243 coal mines in the same 1° by 1° grid. The nearly one-to-one correlation indicates that samples and mines capture the same spatial variability of coal carbon content across China.

¹John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA. ²Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China. ³Resnick Sustainability Institute, California Institute of Technology, Pasadena, California 91125, USA. ⁴Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University, Beijing 100084, China. ⁵School of International Development, University of East Anglia, Norwich NR4 7TJ, UK. ⁶CAS Key Laboratory of Low-carbon Conversion Science and Engineering, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201203, China. ⁷Department of Earth System Science, University of California, Irvine, California 92697, USA. ⁸Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, CE Orme des Merisiers, 91191 Gif sur Yvette Cedex, France. ⁹State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Science, Taiyuan 030001, China. ¹⁰CNRS and UJF Grenoble 1, Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE, UMR5183), 38041 Grenoble, France. ¹¹Department of Geographical Sciences, University of Maryland, College Park, Maryland 20742, USA. ¹²Research Institute for Environment, Energy, and Economics, Appalachian State University, Boone, North Carolina 28608, USA. ¹³Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. ¹⁴Cambridge Centre for Climate Change Mitigation Research, Department of Land Economy, University of Cambridge, 19 Silver Street, Cambridge CB3 9EP, UK. ¹⁵Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China. ¹⁶State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China. ¹⁷Center for International Climate and Environmental Research-Oslo (CICERO), N-0318 Oslo, Norway. ¹⁸CAS Key Laboratory of Pollution Ecology and Environmental Engineering, Chinese Academy of Sciences, Shenyang 110016, China. ¹⁹School of Nature Conservation, Beijing Forestry University, Beijing 10083, China. ²⁰Ecosystems Services & Management Program, International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361 Laxenburg, Austria. ²¹School of Environmental Science and Engineering, South University of Science and Technology of China, Shenzhen 518055, China. ²²State Key Laboratory of Pollution Control & Resource Reuse and School of the Environment, Nanjing University, Nanjing 210023, China. ²³Department of Atmospheric and Oceanic Science and Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20742-2425, USA. ²⁴Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China.

year to year, especially for coal¹⁷. Indeed, using different official sources of activity data and emissions factors can result in estimated emissions that vary by up to 40% in a given year (see Methods).

Here, we present revised estimates of Chinese carbon emissions from the burning of fossil fuels and cement production during the period 1950–2013 using independently assessed activity data and two sets of comprehensive new measurements of emission factors. Results suggest that Chinese CO₂ emissions have been substantially overestimated in recent years: 14% less than the estimates by the Emissions Database for Global Atmospheric Research (EDGAR) version 4.2 (EDGAR being adopted by the Intergovernmental Panel on Climate Change (IPCC) as the emission baseline) in 2013 and 12% less than the latest inventory China reported to the United Nations Framework Convention on Climate Change (UNFCCC; in 2005). The difference is due primarily to the emission factors used to estimate emissions from coal combustion; our measurements indicate that the factors applicable to Chinese coal are on average about 40% lower than the default values recommended by the IPCC^{7,11} and used by previous emissions inventories^{1,4,18}.

In re-evaluating Chinese energy consumption, we adopt the ‘apparent consumption’ approach^{13,15}, which does not depend upon energy consumption data (that previous studies have shown to be not very reliable^{16,19}). Instead, apparent energy consumption is calculated from a mass balance of domestic fuel production, international trade, international fuelling, and changes in stocks, data about which are less subject to ‘adjustment’ by reporting bodies and accounting errors related to either energy consumed during fuel processing or assumptions about the mix of fuel types (especially coal) being used by individual consumers. Furthermore, this approach allows imported and domestically produced fuels to be tracked separately so that appropriate emission factors can be applied to these fuels (see Methods).

Apparent consumption of coal, oil and natural gas in China in 2013 was 3.84 Gt, 401.16 Mt and 131.30 Gm³, respectively. Between 1997 and 2012, we estimate that cumulative energy consumption was 10%

greater than the national statistics and 4% lower than provincial statistics (Extended Data Fig. 2). In addition, our results indicate a higher annual growth rate of energy consumption than national statistics between 2000 and 2010 (9.9% yr⁻¹ instead of 8.8% yr⁻¹); the high growth rate is consistent with satellite observations of NO_x^{20,21}, although NO_x to fuel emission factors change with time as well.

Given the large fraction of CO₂ emissions from coal combustion (80% between 2000 and 2013), estimates of total emissions are heavily dependent on the emission factors used to assess coal emissions. Thus, we re-evaluate each of the variables that determine these emission factors. The mean total carbon content of raw coal samples from 4,243 state-owned Chinese coal mines (4,243 mines represent 36% of Chinese coal production in 2011)²² (Fig. 1) is 58.45% (Fig. 2a), and the production-weighted total carbon content is 53.34%.

These results straddle the results of an independent set of 602 coal samples from the 100 largest coal-mining areas in China (these areas together represent 99% of Chinese coal production in 2011)²² (Extended Data Fig. 3), which reveal a similarly low mean carbon content of 55.48% (Fig. 2b) and a production-weighted mean total carbon content of 54.21%. The net carbon content per energy produced of these same samples is 26.59 tC TJ⁻¹, or 26.32 tC TJ⁻¹ if weighted by production (Fig. 2c), and their net heating value is 20.95 PJ Mt⁻¹, or 20.6 PJ Mt⁻¹ if weighted by production (Fig. 2d). Although the measured net carbon content per energy produced of these samples is within 2% of the IPCC default value (25.8 tC TJ⁻¹), the heating value from these coal samples (20.95 PJ Mt⁻¹) is significantly less than either the IPCC default value (coking coal) of 28.2 PJ Mt⁻¹ or the mean value of US coal of 26.81 PJ Mt⁻¹ (ref. 23). The lower heating value of Chinese coal reflects its generally low quality and high ash content (Fig. 2e, f). For example, the average ash content of our 602 coal samples was 26.91% compared to the average ash content of US coal, 14.08% (ref. 23), but is consistent with recent studies²⁴.

Finally, we assessed the oxidation rate (carbon oxidized per carbon content) of the fossil fuels consumed by 15 major industry sectors in

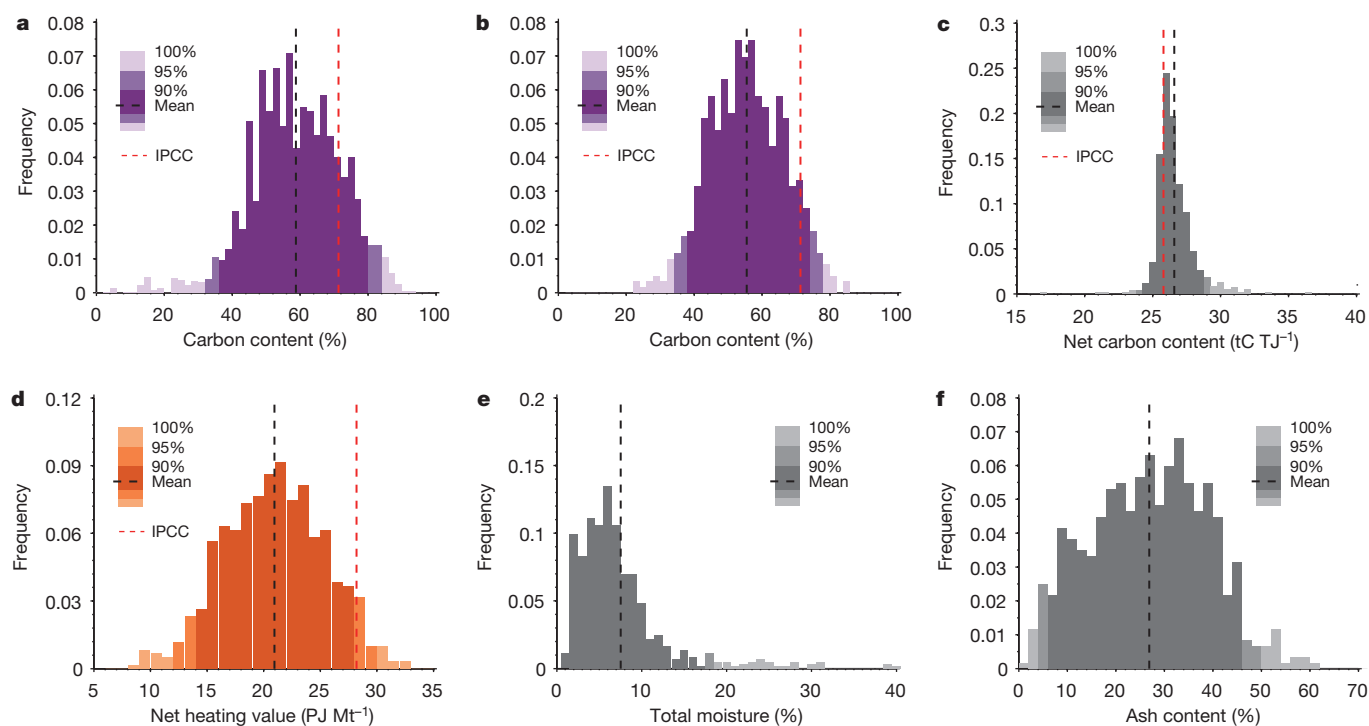


Figure 2 | Histograms of Chinese coal properties. **a, b**, Total carbon content of 4,243 coal mines (**a**) and 602 coal samples (**b**). Dashed lines show mean, and shading indicates 90% and 95% intervals. **c, d**, Net carbon content (**c**) and net heating values (**d**) of the 602 coal samples. Carbon content for coal mines (**a**) and samples (**b**) are significantly lower than IPCC values, which is mainly

because of the lower net heating values of China's coal (**d**); net carbon content is close to the IPCC value (**c**). **e, f**, Total moisture (**e**) and ash content (**f**) further proved the low quality of China's coal, which in general has high ash content but low carbon content.

China with 135 different combustion technologies (see Supplementary Data), as analysed by the National Development and Reform Commission (NDRC; ref. 25). We calculate a production-weighted average oxidation rate for coal of 92%, somewhat lower than the IPCC default value of 98%, but generally consistent with China-specific values reported by the NDRC (94%)²⁵, China's National Communication that reported to UNFCCC (92%)⁸, and the estimate from a previous study (on average 93%)²⁶. Our estimates of the oxidation values of oil and natural gas in China (98% and 99%, respectively) are each within 1% of the IPCC default value.

Combining our revised estimates of carbon content, heating value and oxidation value, we derive new emission factors for coal, natural gas and oil burned in China. The revised emission factors differ from IPCC defaults by -40% , $+13\%$ and -1% , respectively (Fig. 3). In turn, applying these lower emission factors to our revised estimates of energy consumption, our best estimate of Chinese carbon emissions from fossil fuel combustion in 2013 is 2.33 GtC using the carbon content of 4,243 coal mine samples, and 2.31 GtC if the carbon content of 602 coal samples is used. On the basis of the residual scatter of carbon contents from these independent sets of coal samples (Fig. 1), the associated 2σ uncertainty related to coal carbon content

is of the order of 3%. Additional uncertainty about Chinese emissions is provided by varying the estimates of coal consumed, by $\pm 10\%$ as evidenced by the range between national and provincial activity data^{16,19}. Combining these two numbers gives the 7.3% uncertainty range of Chinese fossil fuel carbon dioxide emissions.

We also used clinker production data²⁷ to recalculate CO₂ emissions from cement production (which accounts for roughly 7%–9% of China's total annual emissions in recent years⁴). This direct method avoids use of default clinker-to-cement ratios (for example, 75% and 95% in the IPCC guidelines^{7,11}), and results in emissions estimates that are 32%–45% lower than previous estimates (0.17 GtC in 2012 compared to 0.30 GtC reported by the Carbon Dioxide Information Analysis Center (CDIAC) and 0.24 GtC by EDGAR; Extended Data Fig. 4). The clinker-to-cement ratio calculated by clinker production is 58%, or $\sim 23\%$ lower than the latest IPCC default values. The new, lower estimated cement emissions are consistent with factory-level investigations²⁸ and several other recent studies^{29,30}.

Together, our revised estimate of fossil fuel and cement emissions in 2013 is 2.49 GtC ($2\sigma = \pm 7.3\%$); the new estimate is 12% less than the latest inventories China reported to the UNFCCC (1.63 GtC in 2005, $2\sigma = \pm 8$) and 14% less than the estimate by EDGAR version 4.2 (2.84 GtC in 2013, $2\sigma = \pm 10\%$) (Fig. 4). By *t*-test, our revised estimates of fossil fuel and cement emissions during 2000–2013 are generally lower (at the 90% level) than estimates by EDGAR ($P = 0.016$) and CDIAC ($P = 0.077$).

Our new estimate represents a substantial revision of annual global carbon emissions, reducing the global emissions in 2013 by 0.35 GtC, an amount larger than the reported increase in global emissions between 2012 and 2013 (ref. 31). A systematic reduction of fossil fuel and cement emissions of 0.35 GtC translates into a 15% smaller land sink when this term is calculated as a residual between anthropogenic carbon emissions, atmosphere carbon growth and the ocean carbon sink³¹, and is two times the estimated carbon sink in China's forests (0.18 GtC yr^{-1})⁹. Thus it implies a considerable revision of the global carbon budget³¹. Over the full period 2000 to 2013, the downward

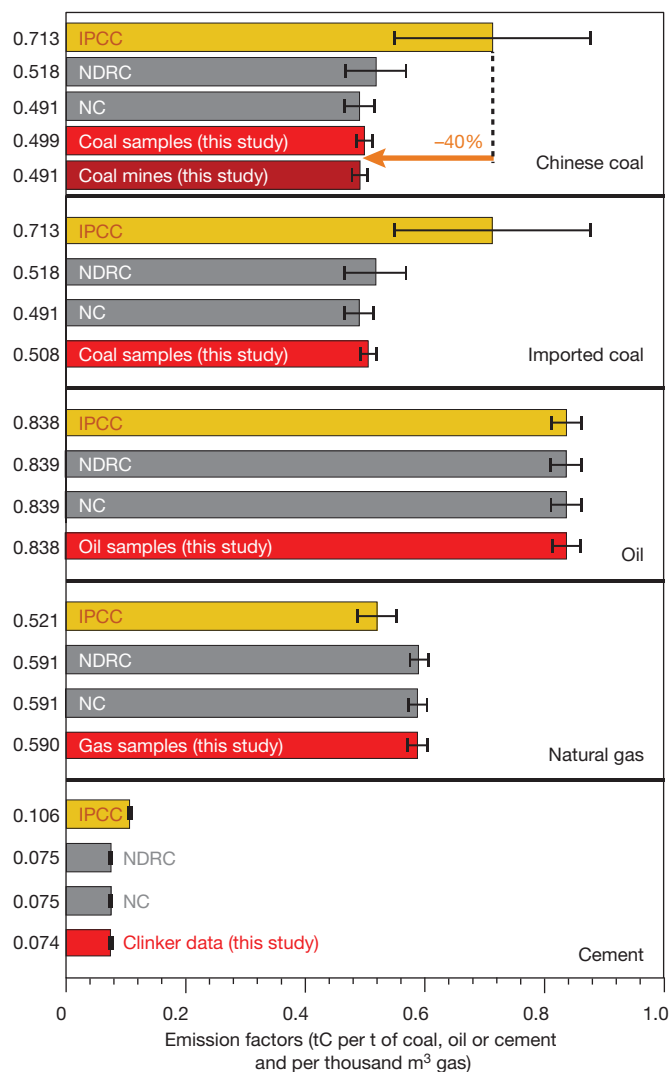


Figure 3 | Comparison of emission factors in 2012. IPCC: default value from IPCC guidelines for national emission inventories (1996, 2006)^{7,11}. NDRC: value reported by the NDRC (ref. 25). NC: China's National Communication, which reported to the UNFCCC (2012 for value in 2005)⁸. All error bars are 2σ errors.

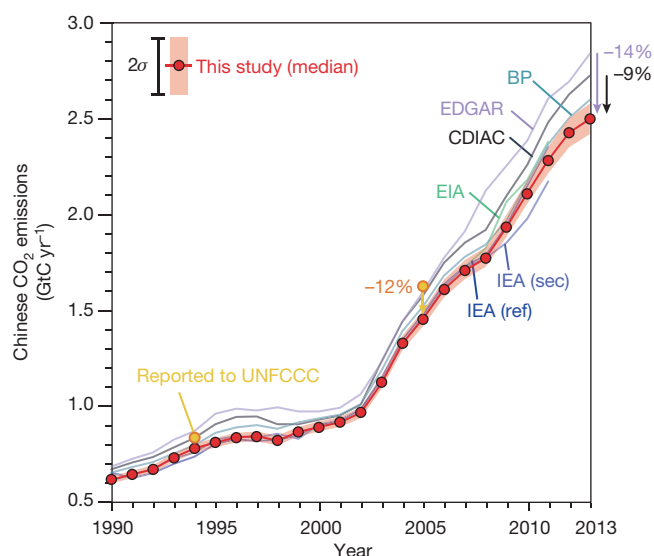


Figure 4 | Estimates of Chinese CO₂ emissions 1990–2013. Total carbon emissions from combustion of fossil fuels and manufacture of cement in China from different sources (International Energy Agency (IEA; sec, sectoral; ref, reference)³, Energy Information Administration (EIA; <http://www.eia.gov/>) and BP¹⁸ estimates do not include the emissions from cement production). The yellow dots are the numbers China reported to the UNFCCC in the years 1994 and 2005 (ref. 8). The red-shaded area indicates the 95% uncertainty range of carbon emissions calculated by this study, assuming the emission factors during the period 1990–2013 are the same as those determined for 2012 in this study.

revision of cumulative emissions in China by 2.9 GtC (13%) is larger than the cumulative forest sink in 1990–2007 (2.66 GtC)⁹ or China's land carbon sink in 2000–2009 (2.6 GtC)¹⁰. Depending upon how the remaining quota of cumulative future carbon emissions is shared among nations, a correction of China's current annual emissions by 10% suggests a 25% (inertia basis) or 70% (blended basis) difference in the cumulative future emissions that can be emitted by China under a 2 °C warming target³². Evaluating progress towards national commitments to reduce CO₂ emissions depends upon improving the accuracy of annual emissions estimates and reducing related uncertainties³³.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 November 2014; accepted 10 June 2015.

1. Boden, T. A., Marland, G. & Andres, R. J. *Global, Regional, and National Fossil-Fuel CO₂ Emissions* (Oak Ridge National Laboratory, US Department of Energy, 2013).
2. Liu, Z. *et al.* A low-carbon road map for China. *Nature* **500**, 143–145 (2013).
3. International Energy Agency. *CO₂ Emissions from Fuel Combustion* (IEA, 2013).
4. Olivier, J. G., Janssens-Maenhout, G. & Peters, J. A. *Trends in Global CO₂ Emissions: 2013 Report* (PBL Netherlands Environmental Assessment Agency, 2013).
5. Kurokawa, J. *et al.* Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in Asia (REAS) version 2. *Atmos. Chem. Phys.* **13**, 11019–11058 (2013).
6. National Bureau of Statistics of China. *Chinese Energy Statistics Yearbook* (China Statistics Press, 2013).
7. Intergovernmental Panel on Climate Change. *2006 IPCC Guidelines for National Greenhouse Gas Inventories* (IPCC, 2006).
8. National Development and Reform Commission. *Second National Communication on Climate Change of the People's Republic of China* (Department of Climate Change, 2012).
9. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
10. Piao, S. *et al.* The carbon balance of terrestrial ecosystems in China. *Nature* **458**, 1009–1013 (2009).
11. Intergovernmental Panel on Climate Change. *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories* (IPCC, 1997).
12. Gregg, J. S., Andres, R. J. & Marland, G. China: emissions pattern of the world leader in CO₂ emissions from fossil fuel consumption and cement production. *Geophys. Res. Lett.* **35**, L08806 (2008).
13. Andres, R. J., Boden, T. A. & Higdon, D. A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. *Tellus B Chem. Phys. Meteorol.* **66**, 23616 (2014).
14. Fridley, D. *Inventory of China's Energy-Related CO₂ Emissions in 2008* (Lawrence Berkeley National Laboratory, 2011).
15. Andres, R. J. *et al.* A synthesis of carbon dioxide emissions from fossil-fuel combustion. *Biogeosciences* **9**, 1845–1871 (2012).
16. Guan, D., Liu, Z., Geng, Y., Lindner, S. & Hubacek, K. The gigatonne gap in China's carbon dioxide inventories. *Nature Climate Change* **2**, 672–675 (2012).
17. Sinton, J. E. & Fridley, D. G. A guide to China's energy statistics. *J. Energ. Lit.* **8**, 22–35 (2002).
18. BP. *BP Statistical Review of World Energy 2014* (BP, 2014).
19. Zhao, Y., Nielsen, C. P. & McElroy, M. B. China's CO₂ emissions estimated from the bottom up: recent trends, spatial distributions, and quantification of uncertainties. *Atmos. Environ.* **59**, 214–223 (2012).
20. Reuter, M. *et al.* Decreasing emissions of NO_x relative to CO₂ in East Asia inferred from satellite observations. *Nature Geosci.* **7**, 792–795 (2014).
21. Lin, J.-T. & McElroy, M. Detection from space of a reduction in anthropogenic emissions of nitrogen oxides during the Chinese economic downturn. *Atmos. Chem. Phys.* **11**, 8171–8188 (2011).
22. National Bureau of Statistics. *China Statistical Yearbook 1996–2014* (China Statistics Press, 2014).
23. Hatch, J. R., Bullock, J. H. & Finkelman, R. B. *Chemical Analyses Of Coal, Coal-Associated Rocks And Coal Combustion Products Collected For The National Coal Quality Inventory* (USGS, 2006).
24. Zhao, Y., Wang, S., Nielsen, C. P., Li, X. & Hao, J. Establishment of a database of emission factors for atmospheric pollutants from Chinese coal-fired power plants. *Atmos. Environ.* **44**, 1515–1523 (2010).
25. National Development and Reform Commission. *Guidelines for China's Provincial GHG Emission Inventories* (NDRC, 2012).
26. Peters, G., Weber, C. & Liu, J. *Construction of Chinese Energy and Emissions Inventory* (NTNU, 2006).
27. China Cement Association. *China Cement Almanac 2012–2013* (China Building Materials Press, 2014).
28. Shen, L. *et al.* Factory-level measurements on CO₂ emission factors of cement production in China. *Renew. Sustain. Energy Rev.* **34**, 337–349 (2014).
29. Liu, M. *et al.* Refined estimate of China's CO₂ emissions in spatiotemporal distributions. *Atmos. Chem. Phys.* **13**, 10873–10882 (2013).
30. Ke, J., McNeil, M., Price, L., Khanna, N. Z. & Zhou, N. Estimation of CO₂ emissions from China's cement production: methodologies and uncertainties. *Energy Policy* **57**, 172–181 (2013).
31. Le Quéré, C. *et al.* Global carbon budget 2014. *Earth Syst. Sci. Data Discuss.* **7**, 521–610 (2014).
32. Raupach, M. R. *et al.* Sharing a quota on cumulative carbon emissions. *Nature Clim. Change* **4**, 873–879 (2014).
33. Liu, Z. *et al.* Climate policy: Steps to China's carbon peak. *Nature* **255**, 279–781 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Strategic Priority Research Program “Climate Change: Carbon Budget and Relevant Issues” of the Chinese Academy of Sciences, and by China's National Basic Research Program and National Natural Science Foundation of China (NSFC) funded projects (grants XDA05010109, 2014CB441301, XDA05010110, XDA05010103, XDA05010101, 41328008 and 41222036). Z.L. acknowledges Harvard University Giorgio Ruffolo fellowship and support from Italy's Ministry for Environment, Land and Sea. D.G. acknowledges the Economic and Social Research Council funded project “Dynamics of Green Growth in European and Chinese Cities” (ES/L016028) and the Philip Leverhulme Prize. S.J.D. acknowledges support from the Institute of Applied Ecology, Chinese Academy of Sciences Fellowships for Young International Distinguished Scientists. P.C. and S.P. acknowledge support of the European Research Council Synergy grant ERC-2013-SyG 610028-IMBALANCE-P. R.J.A. and T.A.B. were sponsored by the US Department of Energy, Office of Science, Biological and Environmental Research under US Department of Energy contract DE-AC05-00OR22725. J. Lin acknowledges the NSFC (41422502 and 41175127). J. Liu acknowledges the International Science & Technology Cooperation Program of China (2012DFA91530), the NSFC (41161140353, 91425303), The Natural Science Foundation of Beijing, China (8151002), the National Program for Support of Top-notch Young Professionals, and Fundamental Research Funds for the Central Universities (TD-JC-2013-2). F.X. acknowledges the NSFC (41473076), China CDM Fund (2013051, 2013124) and Shenyang Science and Technology Planning (F14-232-6-01, F14-134-9-00). G.P.P. acknowledges funding from the Norwegian Research Council (235523). The authors are grateful to S. Piao, L. Cao and J. Yan for insightful comments.

Author Contributions Z.L. and D.G. designed the paper. Z.L. conceived the research. Z.L. provided the data from 4,243 coal mines. W.W. and J.B. provided the measurement data from 602 coal samples. S.J.D., J.B., Q.Z., R.J.A. and T.A.B. provided the reference data. Z.L., D.G., S.J.D., P.C., S.P., J.L., H.Z., C.H., Y.L. and Q.Z. performed the analysis. S.J.D., S.P., Z.L., H.Z. and K.F. drew the figures. All authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.L. (liuzhu@iae.ac.cn), D.G. (dabo.guan@uea.ac.uk), W.W. (weiwei@sari.ac.cn) or K.H. (hekb@tsinghua.edu.cn).

METHODS

Calculation of carbon emissions from fossil fuel combustion and cement production. Carbon emissions are calculated by using activity data, which are expressed as the amount of fossil fuels in physical units used during a production processes (activity data_{clinker} is the amount of clinker produced) multiplied by the respective emission factor (EF):

$$\text{Emission} = \text{activity data} \times \text{EF} \quad (1)$$

Emissions from cement manufacturing are estimated as:

$$\text{Emission}_{\text{cement}} = \text{activity data}_{\text{clinker}} \times \text{EF}_{\text{clinker}} \quad (2)$$

If data on sectorial and fuel-specific activity data and EF are available, total emission can be calculated by:

$$\text{Emission} = \sum \sum \sum (\text{activity data}_{i,j,k} \times \text{EF}_{i,j,k}) \quad (3)$$

where i is an index for fuel types, j for sectors, and k for technology type. Activity data are measured in physical units (tonnes of fuel expressed as t fuel).

EF can be further separated into net heating value of each fuel v , the energy obtained per unit of fuel (TJ per t fuel), carbon content c (tC TJ⁻¹ fuel) and oxidization rate o (in %), the fraction of fuel oxidized during combustion and emitted to the atmosphere). The values of v , c and o are specific for fuel type, sector and technology:

$$\text{Emission} = \sum \sum \sum (\text{activity data}_{i,j,k} \times v_{i,j,k} \times c_{i,j,k} \times o_{i,j,k}) \quad (4)$$

For the coal extracted in China (for example, for the 4,243 coal mines analysed in this study), net heating v and carbon content c values are not directly available, and a more straightforward emission estimate for coal emissions can be obtained using the mass carbon content (C_{ar} in tC per t fuel) of fuels, defined by $C_{\text{ar}} = c \times v$ so that the total emission can be calculated as:

$$\text{Emission} = \sum \sum \sum (\text{activity data}_{i,j,k} \times C_{\text{ar},i,j,k} \times o_{i,j,k}) \quad (5)$$

Apparent energy consumption calculation. The activity data can be directly extracted as the final energy consumption from energy statistics, or estimated based on the mass balance of energy, the so-called apparent energy consumption estimation:

$$\begin{aligned} \text{Apparent energy consumption} = & \text{domestic production} + \text{imports} - \text{exports} \\ & \pm \text{change in stocks} - \text{non-energy use of fuels} \end{aligned} \quad (6)$$

Calculation of carbon emission from cement production. The carbon emission from cement production is due to the production of clinker, which is the major component of cement. When clinker is produced from raw materials, the calcination process of calcium carbonate (CaCO₃) and cement kiln dust (CKD) releases CO₂: CaCO₃ → CaO + CO₂. The amount of emission can be calculated from the molar masses of CaO (55.68 g mole⁻¹) and carbon (12 g mole⁻¹) and the proportion of their masses in clinker production. Furthermore, the emission associated with CKD that is not recycled to the kiln is calculated using the CKD correction factor, CF_{cdk}.

Carbon emission from cement production can be calculated the by clinker emission factor (EF_{clinker}) and clinker production.

$$\text{Emission}_{\text{cement}} = \text{activity data}_{\text{clinker}} \times \text{EF}_{\text{clinker}} \quad (7)$$

$$\text{EF}_{\text{clinker}} = \text{EF}_{\text{CaO}} \times (1 + \text{CF}_{\text{cdk}}) \quad (8)$$

$$\text{EF}_{\text{CaO}_{\text{clinker}}} = \text{fraction CaO} \times (12/55.68) = \text{fraction CaO} \times 0.2155 \quad (9)$$

Fraction CaO is the mass proportion of CaO per unit clinker (in %). EF_{CaO_{clinker}} is the mass of total carbon emission released as CaO per unit of clinker (unit: tC per t clinker). CF_{cdk} is the CKD correction factor (in %). EF_{clinker} is the mass of total carbon emission per unit of clinker (tC per t clinker)

Clinker is the major component of cement. However, data on clinker production is less widely reported than that of cement production. Where data about clinker production are not available, the clinker-to-cement ratio $R_{\text{clinker-cement}}$ (in %) can be used to estimate the cement emission factor (EF_{cement}) and to further estimate the emission based on cement production.

$$R_{\text{clinker-cement}} = \text{activity data}_{\text{clinker}} / \text{activity data}_{\text{cement}} \quad (10)$$

$$\text{EF}_{\text{cement}} = R_{\text{cement-clinker}} \times \text{EF}_{\text{clinker}} \quad (11)$$

$$\text{Emission}_{\text{cement}} = \text{EF}_{\text{cement}} \times \text{activity data}_{\text{cement}} \quad (12)$$

The IPCC default Fraction CaO (clinker) is 64.6%, and the fraction CaO (cement) is 63.5%; thus, the IPCC default EF_{clinker} is 0.1384 (tC per t clinker). In the IPCC 1996 guidelines, the clinker-to-cement ratio is 95%, which assumes that most cement is Portland cement and that the corresponding default EF_{clinker} is 0.1360 (tC per t clinker). In the IPCC 2006 guidelines, the clinker-to-cement ratio is 75% when no direct clinker production data are available, and the corresponding default EF_{clinker} is 0.1065 (tC per t clinker). In this study, the clinker-to-cement ratio is calculated using clinker production statistics and cement production statistics. The cement production and clinker production statistics are listed in Supplementary Information.

It should be noted that the non-energy use of fossil fuels and other industrial processes such as ammonia production, lime production and steel production will also produce carbon emissions. To be consistent with the scope of the international data set we are comparing, those emissions are not included in this study. On the basis of a previous study the total emissions of these non-energy fuel use and industry processes was equivalent to 1.2% of China's emissions from fossil combustion in 2008 (ref. 14).

The uncertainty range of China's emission estimates. We conduct analysis to show the uncertainty range of China's emission estimates based on emission factors (EFs) reported in the literature. We collected 12 sets of EF data for fossil fuel combustion from the six following official sources: IPCC (1996, 2006)^{7,11}, China National Development and Reform Commission (NDRC)³⁴, UN Statistics (UN)³⁵, China National Communication on Climate Change (NC)⁸, China National Bureau of Statistics (NBS)²² and Multi-resolution Emission Inventory for China (MEIC; <http://www.meicmodel.org>). There are three sets of EFs in the NDRC data, corresponding to three tiers of fuel classifications, four sets in NC and two sets in UN. We combined these 12 sets of EFs with two sets of energy statistics derived from national and provincial data^{6,36}. This yielded 24 possible inventories for China's carbon emissions of fossil fuel combustion for 1997–2012 (Extended Data Table 1). The underlying data used in the commonly used data sets (IEA, CDIAC, BP, EDGAR) is either listed in this data assembly (NBS and IPCC) or not publically available.

The mean value of 24 possible inventories is 2,523 MtC in 2012, and the standard deviation is 376 MtC (15%); the detailed data are listed in the Extended Data Table 1. The 2σ standard deviation range suggested by 24 possible inventories is 30%, which is larger than the reported range of 10% by current emission data sets such as EDGAR.

A Monte Carlo (Extended Data Fig. 1) approach was adopted to assess the distribution range of the emissions by assuming that all reported EF values have the same probability (values have been randomly selected with equal probabilities and calculated for 100,000 times). The mean value of the 24 members' ensemble is 2.43 GtC in 2012 (95% confidence interval is +20%, −11% and maximum–minimum range of +27%, −15%). The uncertainty is attributed to the activity data (about 40% of total uncertainty) and EF (60%). The variability of EF for coal dominates the total uncertainty (55% for total uncertainty and 90% for the uncertainty by EF), whereas the EF for other fuels are more comparable. Different EF values for coal mainly reflect variation in v and hence C_{ar} ($C_{\text{ar}} = v \times c$) values, whereas the variation of c and o are comparatively smaller (less than 10%).

The distribution range of the emissions is listed in Extended Data Fig. 1.

We assumed the equal possibility for various EFs when conducting the Monte Carlo analysis; this will expand the uncertainty range. However, both the standard deviation of 24 possible inventories and the Monte Carlo analysis show a significant uncertainty range, implying the considerable system error of the emission estimates by using reported EFs; thus it is critical to perform emission estimates based on a measurement-based EF.

Apparent consumption calculation. We adopted the 'apparent consumption' approach to recalculate China's energy consumption. The apparent energy consumption is the mass balance of fuels produced domestically for energy production, trade, international fuelling and change in stocks:

$$\begin{aligned} \text{Apparent energy consumption} = & \text{domestic production} + \text{imports} - \text{exports} \\ & \pm \text{change in stocks} - \text{non-energy use of fuels} \end{aligned} \quad (13)$$

The calculated apparent energy consumption is usually different from the reported energy consumption in China. For example, our recalculated energy consumption is higher (17% for coal, 2% for oil and 3% for gas) than the national reported energy consumption for 2013.

We believe the resulting estimates of energy consumption to be more accurate than both national and provincial energy statistics, because of the following reasons.

First, national energy statistics may be biased^{16,37,38} because of underreported fuel use in boilers from small factories and workshops^{17,37,38}. In addition, the

adjustment of national statistics by the Chinese government has been discussed in the literature^{39–41}.

Second, provincial energy statistics are also not reliable because of the considerable inconsistencies in provincial aggregated final-consumption energy statistics. When comparing energy consumption with total available energy supply (production plus imports and changes in stocks) in provincial statistics for 2012, coal and oil show differences of 0.25 Gt coal and 81 Mt oil^{6,22}, respectively. In addition, after removing international trade, the amount of exported and imported coal within all provinces should be equal to each other, whereas, in fact, we found an unexplained mismatch of 0.37 Gt coal in provincial aggregated energy statistics, equal to 21% of total domestically traded coal.

Third, the apparent energy consumption is based on production and trade statistics. Chinese data on fuel production and trade statistics are more reliable and consistent than data on final energy consumption. After many years of policies to reduce or close private coal mines, 97% of the coal production in China (3.40 Gt coal in 2011) is from government-owned companies (including central and local governments) that keep good records of the mass of coal extracted^{42,43}. This reliability is supported by the fact that national and provincial statistics on coal production differed by only 10% in 2012 (refs 6, 22), while the same sources reported coal consumption rates that differed by 37% (3.19 Gt for national data versus 4.36 Gt for provincial data). Moreover, coal production and trade data are consistently released earlier than coal consumption data, suggesting that the production data are the original data and therefore less prone to 'adjustment' for political or other purposes. Finally, trade data have also been monitored internationally, so the numbers can be verified by different nations.

Fourth, compared with the final energy consumption approach that involves 20 kinds of primary and secondary energy products, the apparent consumption approach is much simpler: it considers only three primary fuel types (raw coal, crude oil and natural gas) in order to avoid accounting errors due to energy consumed during fuel processing (for example, mass loss in coal washing and coking).

Fifth, the apparent energy consumption approach uses energy production data, which avoids having to deal with uncertain estimations of the mix of different coal types used by each final consumption category. When considering the variation of EFs for different fuel types and sectors, analysis of the sources of uncertainty is more complex. It is difficult to assess specific coal-burning EFs for a myriad of small consumers, and to scale these data up to the national level. Large energy consumers such as power plants continuously mix coal from different sources, which also makes it very difficult to assess national consumption-weighted average EF (weighted by share of different kinds and quality of coal consumed) from a consumption point of view. In contrast, production data can provide the national production-weighted average EF, and thus the national consumption-weighted average EF can be calculated by excluding the coal used for exports, non-energy use and stock changes.

Sixth, the apparent consumption approach allows us to track imported and domestically produced fuels, so that a different EF can be applied.

Between 1997 and 2012, the calculated apparent energy consumption was 10% greater (14% for 2012) than the one reported in national statistics and 4% lower than provincial statistics (Extended Data Fig. 2). The growth rate of apparent energy consumption is consistent with the growth rate of industrial productions (Extended Data Fig. 5).

Sample selection. China's coal resources are mainly concentrated in 100 major coal mine areas from 24 coal mine bases, and there are about 4,000 stable coal mines among these 100 coal mine areas that record coal production. The location of coal sampling is consistent with the distribution of coal mines (Extended Data Fig. 3).

By collecting the coal samples, the following principles are adopted.

First, the sampling spot is based on coal seams under production in one coal mine district, because the properties of coal from within a coal seam are almost the

same. It is guaranteed that at least one sample is collected from each coal seam in one coal mine district.

Second, every coal mine area is sampled, so the 602 samples are across 100 mine areas that cover the majority of the nation's coal production.

Third, there are at least three samples for each coal mine with a production over 5 million tonnes.

Fourth, in the same coal mine district, coal mines with high production are selected preferentially.

Fifth, for sampling within a location, if the samples are collected from a coal pile, they should be collected from at least three different coal piles. If the samples are collected from a conveyor belt, they should be collected three times with several hour intervals from each other. All these three-times-collected samples are merged together and considered as one sample data point (in total 602 sample data points) for further analysis. All samples are stored in sealed plastic bags for further analysis.

Sample analysis. For the sample measurements, we measured the air dry moisture, total moisture, net heating value, and the ash, carbon, hydrogen, nitrogen and total sulfur content. Carbon, hydrogen, nitrogen and total sulfur are determined by combustion using an Elementar elemental analyser. Coal samples are weighed into a tin capsule and burned in a tube furnace at 1,350 °C. Carbon dioxide, water, nitrogen dioxide and sulfur oxide are released from the samples and measured by a thermal conductive detector (TCD). Two parallel samples were tested together each time. The analysis is performed based on ISO standards:

Measurement process (ISO 18283:2006: Hard coal and coke—Manual sampling).

Air dry moisture (ISO 11722:2013: Solid mineral fuels. Hard coal. Determination of moisture in the general analysis test sample by drying in nitrogen).

The total moisture (ISO 589:2008: Hard coal. Determination of total moisture).

Carbon, hydrogen and nitrogen contents (ISO 625:1996: Solid mineral fuels. Determination of carbon and hydrogen. Liebig method; ISO 29541:2010: Solid mineral fuels. Determination of total carbon, hydrogen and nitrogen content. Instrumental method).

Ash content and volatile matter (ISO 11722:2013: Solid mineral fuels. Hard coal. Determination of moisture in the general analysis test sample by drying in nitrogen; ISO 1171:1997: Solid mineral fuels. Determination of ash; and ISO 562:2010: Hard coal and coke. Determination of volatile matter).

The net calorific value (ISO 1928:2009: Solid mineral fuels. Determination of gross calorific value by the bomb calorimetric method and calculation of net calorific value).

Total sulfur contents (ISO 334:2013: Solid mineral fuels. Determination of total sulphur. Eschka method).

34. National Development and Reform Commission. *The People's Republic of China National Greenhouse Gas Inventory* (China Environmental Science Press, 2007).

35. The United Nations. *Energy Statistics Database* (United Nations Publications, 2010).

36. Fridley, E. D. *China Energy Databook—User Guide and Documentation, Version 7.0* (Lawrence Berkeley National Laboratory, 2008).

37. Sinton, J. E. Accuracy and reliability of China's energy statistics. *China Econ. Rev.* **12**, 373–383 (2001).

38. Marland, G. Emissions accounting: China's uncertain CO₂ emissions. *Nature Clim. Change* **2**, 645–646 (2012).

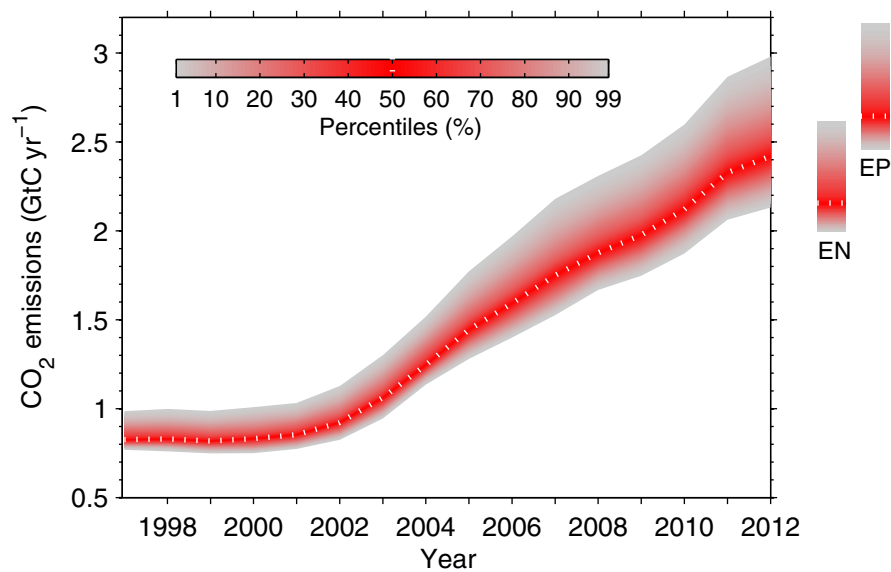
39. Liu, J. & Yang, H. China fights against statistical corruption. *Science* **325**, 675 (2009).

40. Holz, C. A. The quality of China's GDP statistics. *China Econ. Rev.* **30**, 309–338 (2014).

41. Rawski, T. G. What is happening to China's GDP statistics? *China Econ. Rev.* **12**, 347–354 (2001).

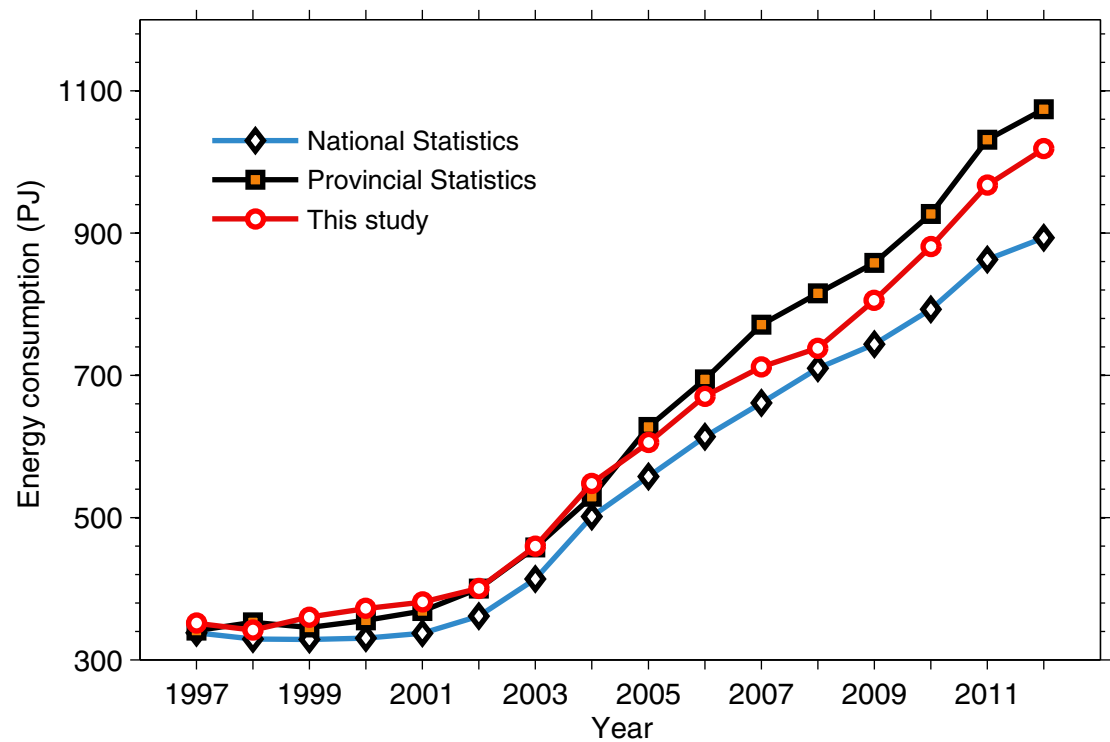
42. Tu, J. *Industrial Organisation of the Chinese Coal Industry* (Freeman Spogli Institute for International Studies, 2011).

43. State Administration of Coal Mine Safety. *China Coal Industry Yearbook* (Coal Information Research Institute, 2013).

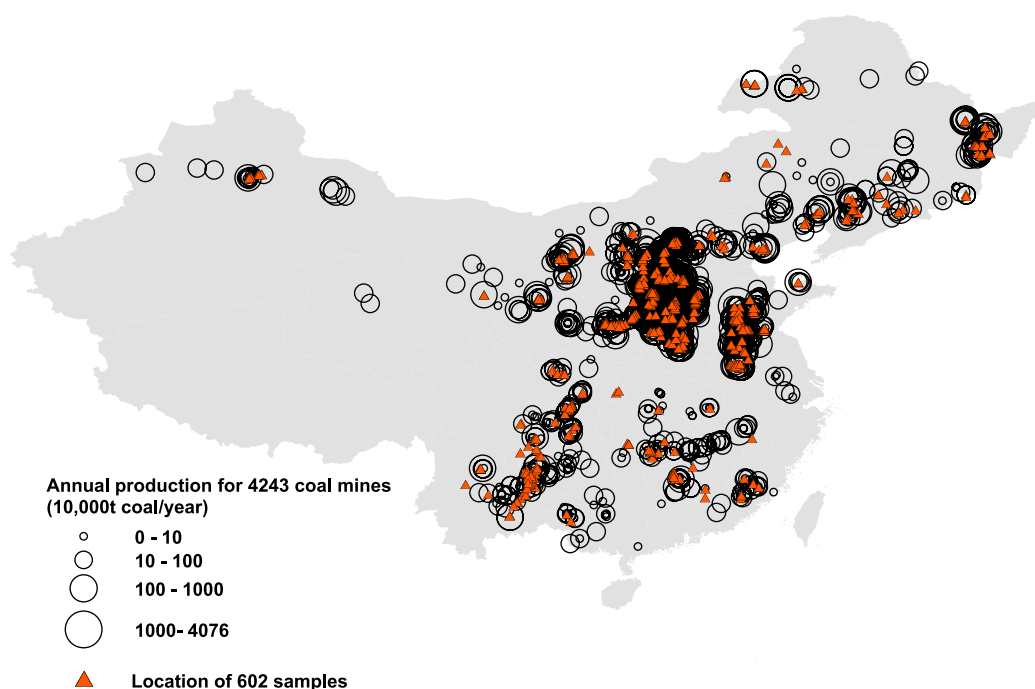


Extended Data Figure 1 | Uncertainty distribution of Chinese CO₂ emissions 1997–2012. Monte Carlo simulations of the Chinese carbon emissions based on a blended activity data set where national and provincial

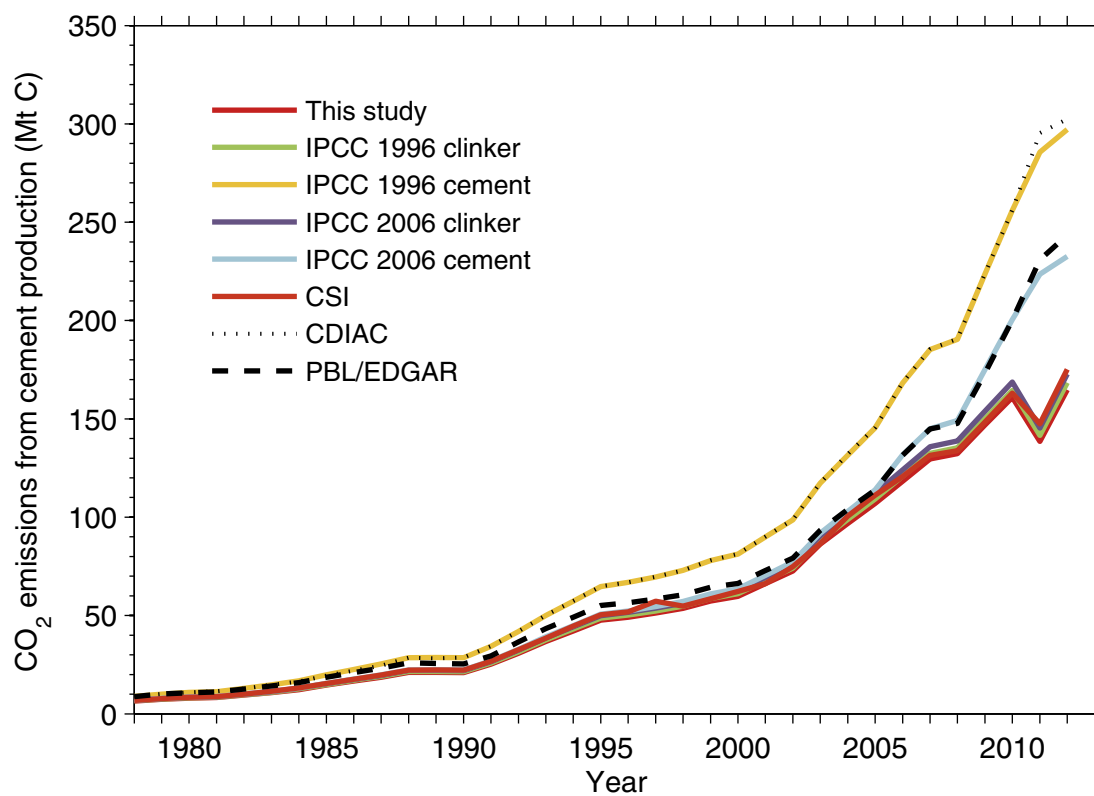
data are assigned equal probabilities ($n = 100,000$). Chinese carbon emissions based on national energy activity data (EN) and provincial activity energy data (EP) in 2012 are shown on the right bar.



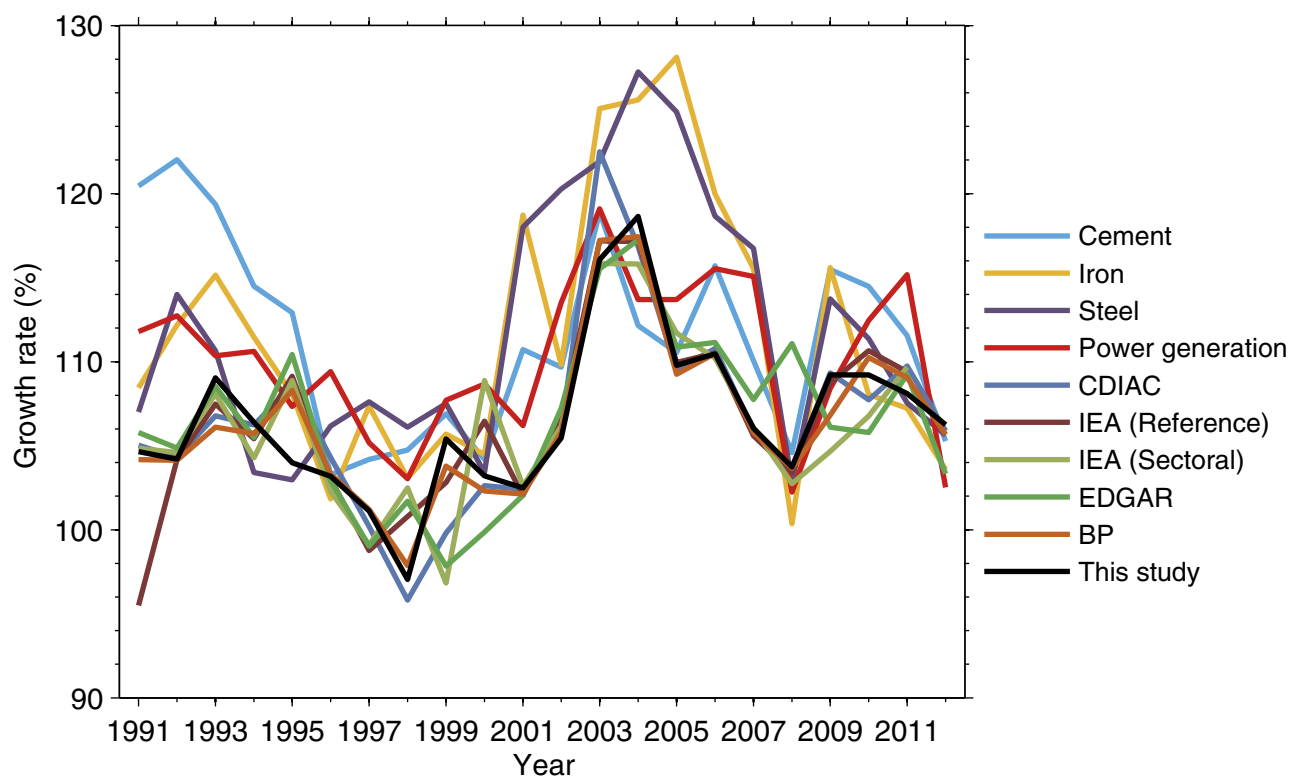
Extended Data Figure 2 | Total fossil fuel energy consumption based upon national statistics, provincial statistics and calculations in this study.



Extended Data Figure 3 | Location of 4,243 coal mines with annual production and 602 coal samples. The coal samples and mines are consistent with spatial distribution.



Extended Data Figure 4 | Emission estimates of China's cement production by different sources.



Extended Data Figure 5 | Growth rate of carbon emissions, based upon BP, EGDAR, IEA and calculations in this study, and industrial products.

Industrial products comprise the production of cement, iron, steel and power

generation. The emission trends calculated in this study are consistent with the trends of industrial production.

Extended Data Table 1 | Twenty-four emission inventories of fossil fuel combustion based on reported emission factors and fuel inventories in China.

Unit: MtC	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
IPCC EP	1081	1113	1087	1109	1153	1256	1438	1662	1969	2180	2432	2573	2707	2964	3251	3379
IPCC EN	1062	1028	1017	1016	1034	1106	1286	1543	1722	1892	2036	2220	2332	2442	2700	2783
NBS EP	837	862	846	868	902	983	1127	1303	1548	1716	1912	2021	2126	2319	2561	2664
NBS EN	827	805	800	803	819	879	1015	1228	1375	1514	1632	1760	1847	1933	2144	2213
NDRC tier 1 EN	840	817	813	816	832	891	1037	1251	1400	1542	1662	1802	1893	1981	2196	2266
NDRC tier 1 EP	851	877	861	881	918	1000	1148	1330	1580	1753	1958	2070	2181	2392	2626	2737
NDRC tier 2 EP	841	868	849	869	903	984	1128	1305	1550	1718	1918	2030	2139	2346	2573	2685
NDRC tier 2 EN	829	805	800	802	817	874	1016	1225	1370	1506	1622	1762	1850	1934	2143	2212
NDRC tier 3 EP	850	879	858	879	910	995	1137	1319	1555	1723	1924	2037	2137	2340	2564	2675
NDRC tier 3 EN	837	816	806	808	821	880	1019	1236	1370	1505	1620	1763	1843	1926	2133	2201
NC 1994 EN	768	750	746	751	766	822	953	1151	1294	1427	1538	1656	1737	1819	2017	2084
NC 1994 EP	774	798	784	805	838	913	1049	1213	1446	1603	1789	1892	1992	2178	2403	2504
NC 2005 tier 1 EN	819	797	790	791	805	860	999	1206	1348	1481	1594	1734	1821	1903	2107	2174
NC 2005 tier 1 EP	831	858	840	859	892	970	1111	1283	1529	1694	1891	2003	2111	2313	2535	2646
NC 2005 tier 2 EN	816	791	784	783	796	849	985	1189	1330	1459	1570	1711	1796	1876	2075	2141
NC 2005 tier 2 EP	828	856	836	854	885	962	1100	1270	1513	1675	1870	1981	2090	2288	2505	2615
NC 2005 tier 3 EN	817	794	785	787	799	856	991	1199	1331	1462	1573	1713	1791	1872	2072	2138
NC 2005 tier 3 EP	829	859	837	857	887	969	1106	1280	1513	1675	1870	1981	2080	2278	2495	2604
MEIC EN	792	770	765	768	783	838	968	1171	1309	1440	1550	1672	1754	1834	2032	2099
MEIC EP	802	829	810	832	862	938	1074	1241	1475	1633	1817	1923	2024	2204	2437	2537
UN average EN	1103	1063	1051	1048	1064	1134	1319	1575	1737	1895	2031	2258	2372	2490	2740	2818
UN average EP	1126	1160	1134	1152	1200	1305	1487	1725	2042	2257	2518	2663	2804	3108	3353	3483
UN China EN	843	814	808	809	822	876	1014	1215	1342	1464	1567	1747	1836	1929	2124	2186
UN China EP	856	883	867	881	921	1001	1140	1328	1577	1745	1949	2061	2175	2430	2596	2707
Standard deviation	105	109	105	107	113	126	142	155	196	219	260	265	280	339	357	376
Average Value	869	870	857	868	893	964	1110	1310	1509	1665	1827	1960	2060	2212	2432	2523

Highly saline fluids from a subducting slab as the source for fluid-rich diamonds

Yaakov Weiss¹, John McNeill², D. Graham Pearson³, Geoff M. Nowell² & Chris J. Ottley²

The infiltration of fluids into continental lithospheric mantle is a key mechanism for controlling abrupt changes in the chemical and physical properties of the lithospheric root^{1,2}, as well as diamond formation³, yet the origin and composition of the fluids involved are still poorly constrained. Such fluids are trapped within diamonds when they form^{4–7} and so diamonds provide a unique means of directly characterizing the fluids that percolate through the deep continental lithospheric mantle. Here we show a clear chemical evolutionary trend, identifying saline fluids as parental to silicic and carbonatitic deep mantle melts, in diamonds from the Northwest Territories, Canada. Fluid–rock interaction along with *in situ* melting cause compositional transitions, as the saline fluids traverse mixed peridotite–eclogite lithosphere. Moreover, the chemistry of the parental saline fluids—especially their strontium isotopic compositions—and the timing of host diamond formation suggest that a subducting Mesozoic plate under western North

America is the source of the fluids. Our results imply a strong association between subduction, mantle metasomatism and fluid-rich diamond formation, emphasizing the importance of subduction-derived fluids in affecting the composition of the deep lithospheric mantle.

Ancient sections of continental lithospheric mantle (CLM) are characterized by multi-stage evolution, involving strong depletion and melt removal followed by variable degrees of ephemeral refertilization^{1,2}. Refertilization, or enrichment, occurs by mantle metasomatism, whereby invading fluids or melts transport mobile components between different mantle reservoirs. This process plays a major part in shaping the mineralogical and geochemical variation in the CLM, as well as in determining its long-term stability, rheology and oxidation state^{1,8}. While many mantle samples reflect the action of metasomatism, including mantle xenoliths and mineral inclusions in diamonds, the nature of the fluids involved can normally only be constrained

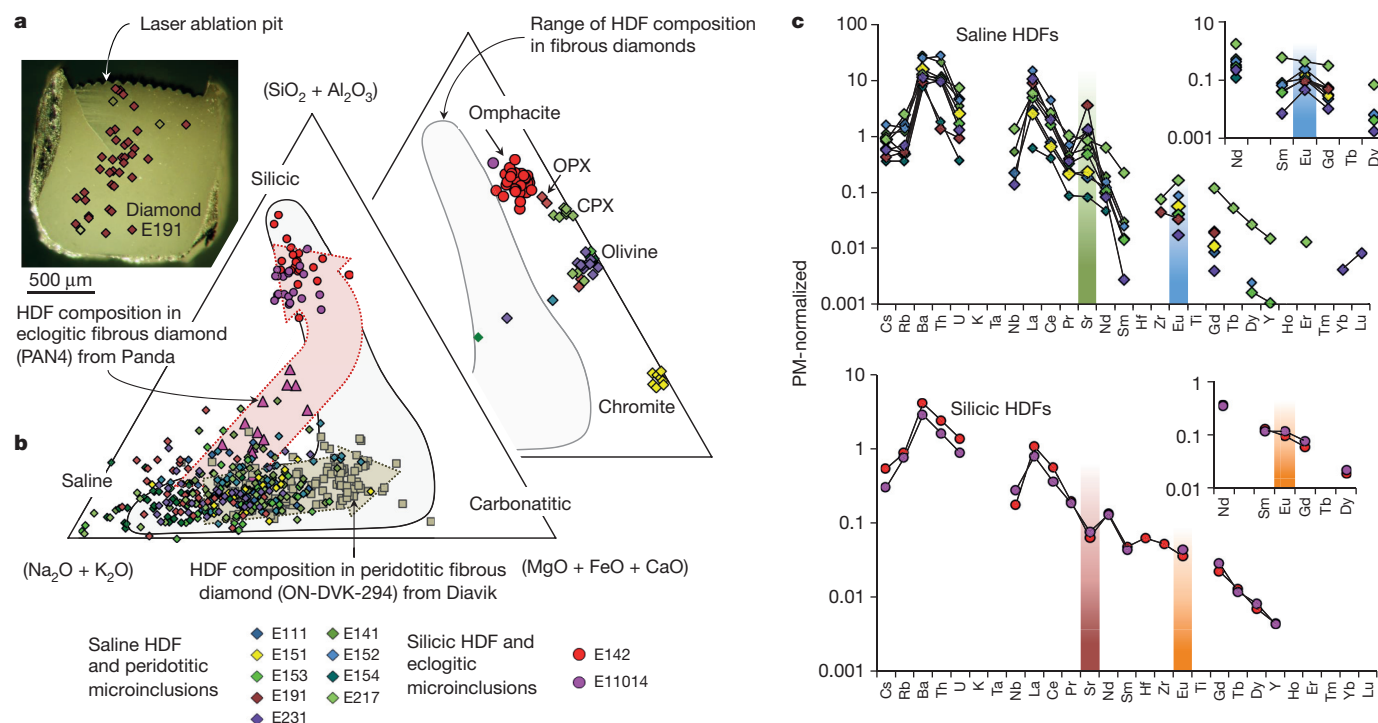


Figure 1 | Microinclusion compositions in fibrous diamonds from the Fox kimberlite, Ekati mine. **a**, Photomicrograph of diamond E191 with the location of the microinclusions analysed by electron probe micro-analyser (EPMA). Filled symbols indicate HDFs; open symbols indicate olivine and orthopyroxene (OPX). **b**, Composition of HDFs and micro-mineral inclusions associated with specific Fox diamonds coded by colour. The global compositional range of HDFs (delineated by average compositions for individual diamonds) and the wide range of compositions shown by individual

diamonds PAN4 (ref. 5) and ON-DVK-294 (ref. 6) from neighbouring central Slave kimberlites are also shown. Shaded arrows define the compositional evolution trajectories of HDFs due to fluid–rock interaction and melting in carbonated peridotite (taupe arrow) and hydrous eclogite (pink arrow) lithologies (see also Fig. 2). **c**, Primitive-mantle (PM) normalized trace element and chondrite-normalized (CN) REE patterns of saline and silicic HDFs in fibrous diamond from the Fox kimberlite. Full analyses and additional figures are in Supplementary Tables 1–4 and Supplementary Fig. 1.

¹Lamont-Doherty Earth Observatory, Columbia University, New York, New York 10964, USA. ²Department of Earth Sciences, Durham University, Durham, DH1 3LE, UK. ³Department of Earth & Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada.

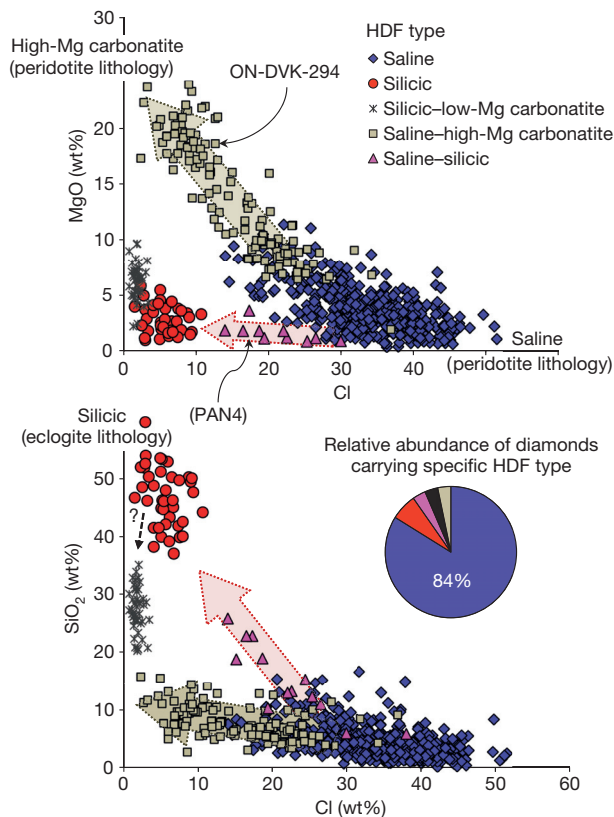


Figure 2 | MgO and SiO₂ versus Cl content of HDF microinclusions in fibrous diamonds from the central Slave craton. The complete data set shows clear evolution trajectories (shaded arrows) from a parental saline fluid to high-Mg carbonatitic and silicic compositions, formed due to wall rock reaction and local melting induced in peridotite and eclogite, respectively. If carbonate is present in the eclogite, increasing reaction could lead to the formation of low-Mg carbonatitic HDFs (dashed black arrow); however, this trend has not yet been constrained by mineral microinclusions paragenesis. Data points for saline HDFs are from this study and refs 5, 6; silicic HDFs from this study; saline to silicic from ref. 5; saline to high-Mg carbonatitic and silicic to low-Mg carbonatitic compositions from ref. 6. Calculated compositions are assumed to be free of H₂O and CO₂ and are in weight per cent (wt%).

indirectly from geochemical proxies or calculated using mineral/melt partition coefficients. Carbonatitic fluid or silicic melts have been proposed as the key metasomatic agents^{9,10}. Direct samples of mantle metasomatic fluids are encased as microinclusions in fast-growing diamonds, known as ‘fibrous diamonds’ (Fig. 1a). These high-density fluid (HDF) inclusions are shielded from late-stage alteration, encapsulating a unique chemical and physical record that can trace the sources of deep mantle fluids and constrain the processes that shape their nature. Their study has revealed that, along with carbonatitic and silicic melts, saline compositions, with very high Cl, K, Na and H₂O contents^{5,6}, are involved in the metasomatic alteration affecting the deepest parts of the CLM.

Diamond HDFs vary between four major compositional types: saline, silicic, and high-Mg and low-Mg carbonatitic⁷ (Fig. 1b). A strong connection was established between high-Mg carbonatitic HDFs and a carbonated peridotite source, either lithospheric or asthenospheric in origin^{7,11,12}, while silicic and low-Mg carbonatitic HDFs have been related to hydrous eclogite (plus or minus carbonate)⁷. The saline fluid endmember sampled by diamonds is more enigmatic and its source in the deep lithosphere has remained ambiguous^{5–7,12,13}.

Here we analysed 11 microinclusion-bearing fibrous diamonds from the Fox kimberlite, Ekati mine, Northwest Territories, Canada. They have either coated or cubic-like morphologies, contain nitrogen in A centres (pairs of nitrogen atoms replacing two adjacent carbon atoms) and encapsulate a variety of fluid compositions plus inclusions

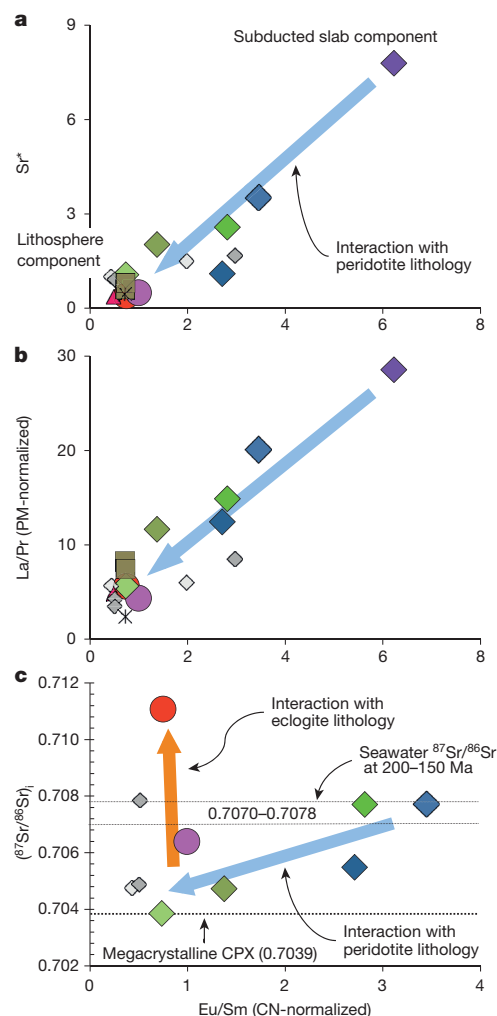


Figure 3 | Trace-element ratios and Sr isotopic signature in HDFs from the central Slave craton. **a**, Relationship between Eu* (expressed here as Eu/Sm) and Sr* (Sr/(Pr × Nd)) anomalies in the saline HDFs constrain the subducted endmember, least influenced by interaction with lithosphere wall rock. **b**, Eu* versus La/Pr ratios. **c**, Eu* versus (⁸⁷Sr/⁸⁶Sr)_i (corrected to kimberlite eruption age of 55 Ma). The positive trend formed by saline fluids varies between the Sr isotopic signature of sea water 150–200 Ma (ref. 26) and ⁸⁷Sr/⁸⁶Sr measured in megacrystalline clinopyroxene (CPX) from the Diavik CLM²¹. Higher (⁸⁷Sr/⁸⁶Sr)_i values in silicic HDFs are probably inherited from old phlogopite in the eclogitic lithology within the Slave CLM. HDF composition symbols as in Fig. 2, large symbols are data from the present study, where each colour represents an individual diamond; small symbols are data from refs 5, 6, 12 and 14.

of their host rocks (Fig. 1a, b and Extended Data Figs 1, 2), which shows a strong association between fluid composition and mantle host lithology. The majority of diamonds (9 of 11) contain saline HDFs solely associated with peridotite on the basis of their microinclusions of olivine, orthopyroxene, Cr-diopside and chromite. Silicic fluid compositions are related exclusively to eclogitic inclusions of omphacitic clinopyroxene. Both saline and silicic HDFs are enriched in incompatible elements (Fig. 1c); they have fractionated rare-earth element (REE) patterns with elevated Ba, U, Th and light-REEs but depleted Nb, Ta and alkalis (K, Rb and Cs). However, the fractionated nature of these patterns and the light-REE/medium-REE and Th/U ratios in particular are more pronounced in saline HDFs than silicic fluids, indicating different sources. The most striking differences between the two HDF compositions are the positive Eu and Sr anomalies within saline fluids versus no Eu anomaly and negative Sr anomalies in the silicic fluids. Initial ⁸⁷Sr/⁸⁶Sr ((⁸⁷Sr/⁸⁶Sr)_i) values in

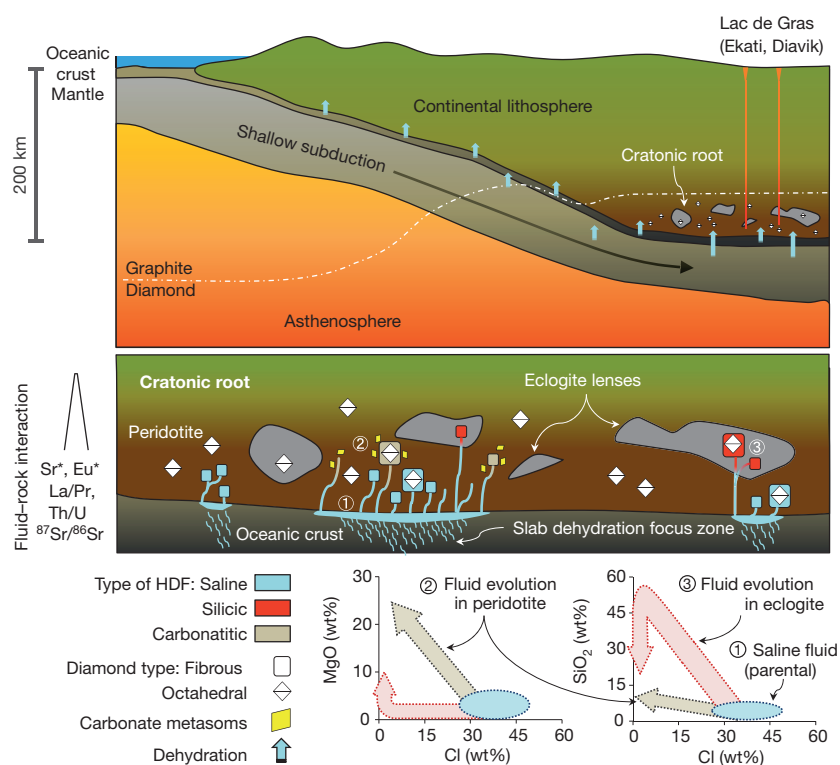


Figure 4 | Schematic illustrating the evolution of saline fluids with increasing fluid–rock interaction as they interact with cratonic mantle lithosphere. The discovery of fluids carrying strong oceanic protolith geochemical signatures (that is, positive Eu and Sr anomalies and $^{87}\text{Sr}/^{86}\text{Sr}$ signature of Mesozoic sea water) in continental lithospheric diamonds suggests that the Slave CLM was directly overlying the subducting slab at the time of Mesozoic metasomatism. Numbers refer to stages in the compositional evolution depicted in the inset MgO–Cl and SiO₂–Cl trends. When the parental saline fluids (1) ingress into the CLM they react and their oceanic signature is diluted. The melt-depleted nature of cratonic lithospheric peridotite prevents notable melting unless the saline fluids traverse either carbonated-peridotite or eclogite lenses, leading to *in situ* formation of high-Mg carbonatitic (2) and silicic melts (3), respectively. The possible presence of carbonate in eclogite may lead to formation of low-Mg carbonatite fluids with increasing melting (see Fig. 2 and Extended Data Fig. 5 for data). Rapid diamond formation occurs due to the oxidation gradient between the evolving fluids and local lithosphere, either as new fibrous diamonds or as fibrous coats on previously formed older octahedral diamonds.

the saline HDFs are 0.7039–0.7090 compared with 0.7064 and 0.7111 in the two diamonds with silicic fluids.

The physical and chemical characteristics of fibrous diamonds and their HDFs sampled by the Fox kimberlite are identical to previously studied fibrous diamonds from both the neighbouring (8 km north-east) Panda kimberlite at Ekati^{5,14} and to those from kimberlites at Diavik mine^{6,12} (30 km southeast). Combining these localities reveals that the vast majority of these fibrous diamonds (84%) trapped saline HDFs, which are strongly associated with peridotite hosts. In a single Diavik diamond, the HDFs continuously change from saline to high-Mg carbonatitic compositions, from centre to edge (Figs 1b, 2 and Extended Data Fig. 3); olivine, chromite and Cr-diopside microinclusions in this diamond demonstrate its peridotitic association⁶. A Panda diamond containing HDFs falling between saline to silicic compositions has omphacite microinclusions⁵, providing strong evidence that silicic HDFs may evolve from saline fluids due to wall rock reaction with eclogite (Figs 1b and 2). An absence of included minerals prevented paragenetic typing in only one diamond, from Diavik, containing silicic to low-Mg carbonatitic HDFs (Fig. 2). However, the continuous global compositional array between silicic to low-Mg carbonatitic HDFs, and the similarity of these fluids to the products of low-degree partial melting experiments of carbonated eclogite, suggest a strong genetic link to eclogite⁷. The relative abundance of HDF endmembers in fibrous diamonds, the compositional relationships between the HDFs and their co-existing mineral microinclusions, plus the observed evolutionary trends from saline HDFs to other compositional types (Fig. 2), provide a means of tying the various metasomatic fluids to a common parental saline fluid endmember. A key issue is then the ultimate origin for saline HDFs.

Positive Eu anomalies in low-pressure Cl[−]-rich hydrothermal fluids are typically interpreted to result from plagioclase control during fluid–rock interaction at high temperatures¹⁵. Such signatures can also originate from the strong aqueous complexes formed between dissolved Cl[−] and Eu²⁺, compared to other REE ions¹⁶. The lack of clear correlation between Cl content and the size of the Eu anomaly in global saline HDFs precludes a simple fluid–rock interaction process being the sole driver for generating the positive Eu anomalies. In Ekati and

Diavik diamonds, the pronounced Eu anomalies of saline HDFs are associated with positive Sr anomalies (Figs 1c and 3a), mimicking the plagioclase accumulation signature of both oceanic and ophiolitic gabbros^{17,18}. This correlation suggests a low-pressure crustal origin for the saline HDF elemental signature, through prograde metamorphic reaction of plagioclase to garnet and eclogite formation. Exactly how the saline chemistry of these fluids develops within subducting oceanic crust is not yet clear. One possibility is that they were originally pore fluids trapped in the crust during low-pressure hydrothermal alteration by sea water. Their initial salinity and K/Na ratios evolve as H₂O and Na are consumed during spilitization, producing hydrated basalt. The highly saline nature of these new solutions potentially prevents dehydration during shallow subduction, allowing the formation of stable Cl[−]-rich phengite at high pressure and temperature^{19,20}. If dehydration should occur, residual chlorides with high K/Na ratios can be subducted and water originating from dehydration of underlying serpentinized peridotite at 150–200 km depth¹⁹ can regenerate highly saline fluids at depth.

Having established a possible evolutionary link between the spectrum of fluid compositions observed in Slave diamonds and subduction-related crustal protoliths, we can deduce the metasomatic history of the central Slave lithospheric root leading to fibrous diamond formation (Fig. 4). The inherited positive Eu and Sr anomalies in saline diamond HDFs suggest direct ingress of fluids into the lithosphere from a subducting slab closely underlying the continental root. We explain the Eu and Sr anomalies, light-REE/medium-REE enrichment levels and variably radiogenic $^{87}\text{Sr}/^{86}\text{Sr}$ in these fluids (Fig. 3a–c) as representing the interaction with peridotite in the lithospheric root. This interaction altered the elemental chemistry of the invading saline fluids, flattening both the Sr and Eu anomalies and lowering the ratios of the most incompatible elements. The $^{87}\text{Sr}/^{86}\text{Sr}$ signature of the saline HDFs experiencing the most extensive fluid–rock interaction were buffered to local peridotite compositions that were relatively low (~0.704) (ref. 21). Fluid–rock interaction is also reflected by an increase of SiO₂, MgO and CaO and decrease of Cl and K in the saline HDFs (Figs 1b and 2). The refractory nature of cratonic peridotite dictated that partial melting during saline fluid infiltration only occurred when carbonate metasomes (that is, magnesite) were

intersected, leading to the formation of high-Mg carbonatite HDFs with low $^{87}\text{Sr}/^{86}\text{Sr}$ (ref. 12). Infiltration of saline fluids into eclogite hosts is tracked by the compositional variation in eclogite-related HDFs, from saline to highly silicic (Figs 1b and 2), leading to the formation of *in situ* silicic melts. Both of the daughter high-Mg carbonatitic and silicic melts could then crystallize metasomatic phases in their host rocks such as the Cl^- -rich phlogopite and apatite documented in an eclogitic xenolith from Diavik²². Overall, the Slave CLM was enriched with K, Cl, Ba and incompatible trace elements by the invading saline HDFs, while the oxidation gradient between the evolving fluids and the local lithosphere initiated ephemeral redox processes leading to diamond formation²³.

The issue is the timing of the fluid metasomatism and the nature of the event that triggered the process. The short mantle residence time of fibrous diamonds in the central Slave CLM (<200 million years ago (Ma); Extended Data Fig. 4), indicated by their low-aggregated nitrogen impurities, translates into young formation ages for both diamonds and their HDFs. Active subduction zones were a key feature of the complex tectonic setting of western North America and the high Arctic during the Mesozoic era²⁴, providing several options for the fluid source in the ideal time window to allow saline HDF generation, diamond formation and eruption of the diamonds in Eocene epoch kimberlites. The low-angle subduction that has been suggested for some of these plates, such as the Farallon slab²⁵, provides an opportunity for the direct transfer of slab-derived fluids into the base of the cratonic lithosphere. The most pristine saline HDFs that interacted least with the lithospheric root have Sr isotopic signatures corresponding with early Jurassic period seawater $^{87}\text{Sr}/^{86}\text{Sr}$ values²⁶, strengthening the temporal connection between subduction and metasomatism (Fig. 3c). The lithosphere beneath western North America was extensively hydrated by shallow subduction^{27,28} and mantle xenoliths from the Wyoming craton provide direct evidence for chlorine enrichment^{29,30}. Saline HDFs trapped in fibrous diamonds from the central Slave craton are a deeper manifestation of this lithospheric hydration process, expressed as young diamond formation in the CLM root. The full spectrum of HDF compositional varieties, from saline, silicic and carbonatitic, are present in fibrous diamonds from various cratonic roots^{6,7,12,13}, including intermediate composition between saline to silicic found in an eclogitic zoned diamond from Guinea⁷ (Extended Data Fig. 5). We suggest that deep mantle saline fluids are directly related to subduction events—they are key metasomatic agents from which the whole spectrum of diamond-forming fluids evolve, and they play a major part in impacting the composition of the deep lithospheric mantle globally.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 February; accepted 19 June 2015.

- Carlson, R. W., Pearson, D. G. & James, D. E. Physical, chemical and chronologic characteristics of continental mantle. *Rev. Geophys.* **43**, RG1001 (2005).
- Griffin, W. L., O'Reilly, S. Y., Afonso, J. C. & Begg, G. The composition and evolution of lithospheric mantle: a re-evaluation and its tectonic implications. *J. Petrol.* **50**, 1185–1204 (2009).
- Shirey, S. B. *et al.* Diamonds and the geology of mantle carbon. *Rev. Mineral. Geochem.* **75**, 355–421 (2013).
- Navon, O., Hutcheon, I. D., Rossman, G. R. & Wasserburg, G. J. Mantle derived fluids in diamond micro inclusions. *Nature* **335**, 784–789 (1988).
- Tomlinson, E. L., Jones, A. P. & Harris, J. W. Co-existing fluid and silicate inclusions in mantle diamond. *Earth Planet. Sci. Lett.* **250**, 581–595 (2006).
- Klein-BenDavid, O., Izraeli, E. S., Hauri, E. & Navon, O. Fluid inclusions in diamonds from the Diavik mine, Canada and the evolution of diamond-forming fluids. *Geochim. Cosmochim. Acta* **71**, 723–744 (2007).
- Weiss, Y. *et al.* A new model for the evolution of diamond-forming fluids: evidence from microinclusion-bearing diamonds from Kankan, Guinea. *Lithos* **112**, 660–674 (2009).
- Creighton, S., Stachel, T., Eichenberg, D. & Luth, R. W. Oxidation state of the lithospheric mantle beneath Diavik diamond mine, central Slave craton, NWT, Canada. *Contrib. Mineral. Petrol.* **159**, 645–657 (2010).
- Rudnick, R. L., McDonough, W. F. & Orpin, A. in *Kimberlites, Related Rocks and Mantle Xenoliths* Vol. 1 (eds Meyer, H. O. A. & Leonardos, O.) 336–353 (Proc. Fifth Int. Kimb. Conf., C.P.R.M., 1994).

- Stachel, T. *et al.* The trace element composition of silicate inclusions in diamonds: a review. *Lithos* **77**, 1–19 (2004).
- Weiss, Y., Griffin, W. L., Bell, D. R. & Navon, O. High-Mg carbonatitic melts in diamonds, kimberlites and the subcontinental lithosphere. *Earth Planet. Sci. Lett.* **309**, 337–347 (2011).
- Klein-BenDavid, O. *et al.* The source and time-integrated evolution of diamond-forming fluids—trace element and Sr isotopic evidence. *Geochim. Cosmochim. Acta* **125**, 146–169 (2014).
- Weiss, Y., Griffin, W. L. & Navon, O. Diamond-forming fluids in fibrous diamonds: the trace-element perspective. *Earth Planet. Sci. Lett.* **376**, 110–125 (2013).
- Tomlinson, E. L., Müller, W. & EIMF. A snapshot of mantle metasomatism: trace element analysis of coexisting fluid (LAICP-MS) and silicate (SIMS) inclusions in fibrous diamonds. *Earth Planet. Sci. Lett.* **279**, 362–372 (2009).
- Klinkhammer, G. P., Elderfield, H., Edmond, J. & Mitra, A. Geochemical implications of rare earth element patterns in hydrothermal fluids from mid-ocean ridges. *Geochim. Cosmochim. Acta* **58**, 5105–5113 (1994).
- Allen, D. E. & Seyfried, W. E. REE controls in ultramafic hosted MOR hydrothermal systems: an experimental study at elevated temperature and pressure. *Geochim. Cosmochim. Acta* **69**, 675–683 (2005).
- Bach, W. *et al.* The geochemical consequences of late-stage low-grade alteration of lower ocean crust at the SW Indian Ridge: results from ODP Hole 735B (Leg 176). *Geochim. Cosmochim. Acta* **65**, 3267–3287 (2001).
- Benoit, M., Polve, M. & Ceuleneer, G. Trace element and isotopic characterization of mafic cumulates in a fossil mantle diapir (Oman ophiolite). *Chem. Geol.* **134**, 199–214 (1996).
- Schmidt, M. W. & Poli, S. Generation of mobile components during subduction of oceanic crust. *Treatise Geochem.* **3**, 567–591 (2003).
- Nazzari, S. *et al.* Synthetic hypersilicic Cl bearing mica in the phlogopite-celadonite join: a multimethodical characterization of the missing link between di- and tri-octahedral micas at high pressures. *Am. Mineral.* **93**, 1429–1436 (2008).
- van Acherbergh, E. *et al.* Subduction signature for quenched carbonates from the deep lithosphere. *Geology* **30**, 743–746 (2002).
- Aulbach, S., Pearson, N. J., O'Reilly, S. Y. & Doyle, B. J. Origins of xenolithic eclogites and pyroxenites from the central Slave craton, Canada. *J. Petrol.* **48**, 1843–1873 (2007).
- Jacob, D. E., Dobrzynetskaya, L. & Wirth, R. New insight into polycrystalline diamond genesis from modern nanoanalytical techniques. *Earth Sci. Rev.* **136**, 21–35 (2014).
- Shephard, G. *et al.* Circum-Arctic mantle structure and long-wavelength topography since the Jurassic. *J. Geophys. Res. Solid Earth* **119**, 7889–7908 (2014).
- Currie, C. A. & Beaumont, C. Are diamond-bearing Cretaceous kimberlites related to low-angle subduction beneath western North America? *Earth Planet. Sci. Lett.* **303**, 59–70 (2011).
- Jones, C. E. & Jenkyns, H. C. Seawater strontium isotopes, oceanic anoxic events, and seafloor hydrothermal activity in the Jurassic and Cretaceous. *Am. J. Sci.* **301**, 112–149 (2001).
- Dixon, J. E., Dixon, T. H., Bell, D. R. & Malservisi, R. Lateral variation in upper mantle viscosity: role of water. *Earth Planet. Sci. Lett.* **222**, 451–467 (2004).
- Smith, D. *et al.* Evolution of Navajo eclogites and hydration of the mantle wedge below the Colorado Plateau, southwestern United States. *Geochim. Geophys. Geosyst.* **5**, (2004).
- Facer, J., Downes, H. & Beard, A. *In situ* serpentinization and hydrous fluid metasomatism in spinel dunite xenoliths from the Bearpaw Mountains, Montana, USA. *J. Petrol.* **50**, 1443–1475 (2009).
- Rowe, M. C. & Lassiter, J. C. Chlorine enrichment in central Rio Grande Rift basaltic melt inclusions: evidence for subduction modification of the lithospheric mantle. *Geology* **37**, 439–442 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements Y.W. acknowledges his Lamont postdoctoral fellowship and National Science Foundation grant no. 1348045. We thank T. Stachel and D. Walker for discussions and J. J. Gurney, J. Carlson, T. Nowicki and BHP Minerals/Dominion Diamonds for access to diamonds from the Ekati mine. J.M. was funded by a scholarship from the Diamond Trading Company at Durham University. D.G.P. completed this work under tenure of a Canada Excellence Research Chair, with support from the Deep Carbon Observatory (Sloan Foundation). Y.W. thanks Israel Science Foundation grant number 435/12 for funding the EPMA and Fourier transform infrared (FTIR) analyses at the Hebrew University. D. E. Jacob, M. Santosh and M. Walter made excellent suggestions that greatly improved this paper. This is Lamont-Doherty Earth Observatory contribution number 7908.

Author Contributions Y.W. performed the EPMA and FTIR analyses and conceived and developed the model. Y.W. and D.G.P. wrote the paper. D.G.P., G.M.N. and J.M. jointly developed the *in situ* closed-cell laser ablation analytical technique used. D.G.P. supervised the trace element and isotopic measurements performed by J.M. C.J.O. aided in sample preparation and measurements. All authors contributed intellectually to the paper.

Author Information Sample metadata have been archived in the System for Earth Sample Registration (SESAR) with associated International GeoSample Numbers (IGSNs). The data set can be found in the EarthChem library (<http://doi.org/10.1594/IEDA/100540>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.W. (yweiss@ldeo.columbia.edu).

METHODS

Samples and methods. A suite of eleven diamonds from the Ekati mine, Slave Craton, Canada, was selected for EPMA, Fourier-transform infrared (FTIR) and off-line laser ablation inductively coupled plasma mass spectrometry (ICP-MS) analyses. The diamonds have a large range in size with weight varying between 3–83 mg. Each diamond was laser-cut and polished on both sides to create a thin slab that permits the transmittance of light. It was then cleaned ultrasonically in HF 60% and HNO₃ 69% for 2 h and washed with ethanol and distilled water before analysis.

FTIR. Analyses were performed using a Bruker IFScope II microscope coupled to a Nicolet 740 FTIR spectrometer (Global source, KBr beamsplitter, MCT detector, He–Ne laser). Spectra were taken in the range of 550–4,000 cm^{−1} with a resolution of 4 cm^{−1}. Nitrogen concentration and aggregation states were determined using a computer program supplied by D. Fisher and the absorption coefficients of A centres (double substitution of carbon by two nitrogen atoms, type IaA spectrum), B centres (clusters of four nitrogen atoms substituting five carbon atoms, type IaB spectrum) and C centres (single nitrogen replacing a carbon atom, type Ib spectrum)^{31–34}. After baseline correction and subtraction of the diamond bands, the concentrations of water and carbonate were determined using the maximum absorbance of water and carbonate and their absorption coefficients³⁵. These concentrations were used to calculate the carbonate mole fraction (CMF = carbonate/(water + carbonate) molar ratio) of the trapped fluids (Supplementary Table 1).

EPMA. The major element compositions of the microinclusions were determined using a JEOL JXA 8600 EPMA equipped with a Pioneer-Norvar EDS (133 eV) detector. Backscattered electron imaging was used to detect shallow, subsurface microinclusions (<2 µm depth). Each inclusion was analysed for 100 s using an acceleration voltage of 15 kV and a beam current of 10 nA. The spectral data were reduced using the ZAF/PROZA correction procedure software supplied by Noran³⁶. The total amount of oxides and Cl in each analysis varied between 1 and 12.4 wt% with an average of 3.3 wt% for all 327 analysed HDF microinclusions and between 1.8 and 78 wt% with an average of 11 wt% for 68 analysed mineral microinclusions. Precision (2σ (%)) = 2 × 1/oxide in wt%) is <20% for oxide concentrations of 0.05 wt%, <10% for 0.25 wt%, <6% for 0.5 wt% and <2% for 1 wt% (M. Jablon and O. Navon, unpublished data). The low and variable sums reflect the small size of the inclusions, their depth and their high content of undetected water and carbonates. The ZAF/PROZA processing assumed that the difference to 100 wt% is composed of pure carbon. Later, all oxide and chlorine concentrations were normalized to 100 wt% on a carbon-free and volatiles-free basis (where Cl is present, excess calculated oxygen leads to a normalized total of more than 100%) and the average composition of the HDF in the diamond was calculated.

Offline laser ablation. Diamonds were ablated in a custom-designed, sealed PTFE ablation cell capped with a laser window that had been previously cleaned with UpA 6 N HCl and 2 N HNO₃. Ablations were performed with a UV-213 New-Wave Laser ablation system, with the custom cell replacing that provided by the manufacturer. A pre-weighed diamond was brought into focus and an ablation was performed using a raster pattern. Ablation conditions were: scan speed 50 µm s^{−1}; raster spacing 80 µm; energy output 5–6 J cm^{−2}; repetition rate 20 Hz; spot size 160 µm; and pass depth 2 µm. Ablation time varied from 3–5 h. After ablation, the laser cell is opened in an ultraclean environment and all ablated material was collected in UpA 6 N HCl before being dried down before further chemistry. The diamond was rinsed in MQ water and dried. Diamonds were re-weighed and the weight loss (0.32–0.71 mg) resulting from the ablation was calculated. Weighing uncertainty is ±0.0007 mg, estimated from 100 repeat weighs of both a gem-quality and a fibrous diamond. The dried ablation product was taken up in 2 N HNO₃. A 20% aliquot is taken by volume for trace element analysis. The remaining sample was processed for Sr isotopic analysis. The Sr separation procedure is based on the method described previously³⁷, using Sr-spec resin but with modifications as outlined³⁸ for sub-ng samples.

Quantifiable data and background corrections. We use the limit of quantification (LOQ), as defined previously³⁹, as a measure of our ability to quantitatively measure elemental abundances because this parameter is considerably more robust than defining the ‘limits of detection’, or LOD, which merely define the ability to qualitatively detect an analyte. The LOQ for a procedure with a well-characterized blank is defined³⁹ as: LOQ = 10σ, where σ is the standard deviation of the blank for the process (here defined as the total procedural blank (TPB)). This approach places clear limits on our ability to quantitatively report concentration data in the diamonds studied. We use a data set of 20 TPBs performed using the same ablation cells and reagents as used for samples, to determine the LOQ for trace element abundances. Within each batch of samples, between five and ten additional TPBs were also run to monitor whether our LOQ estimate was

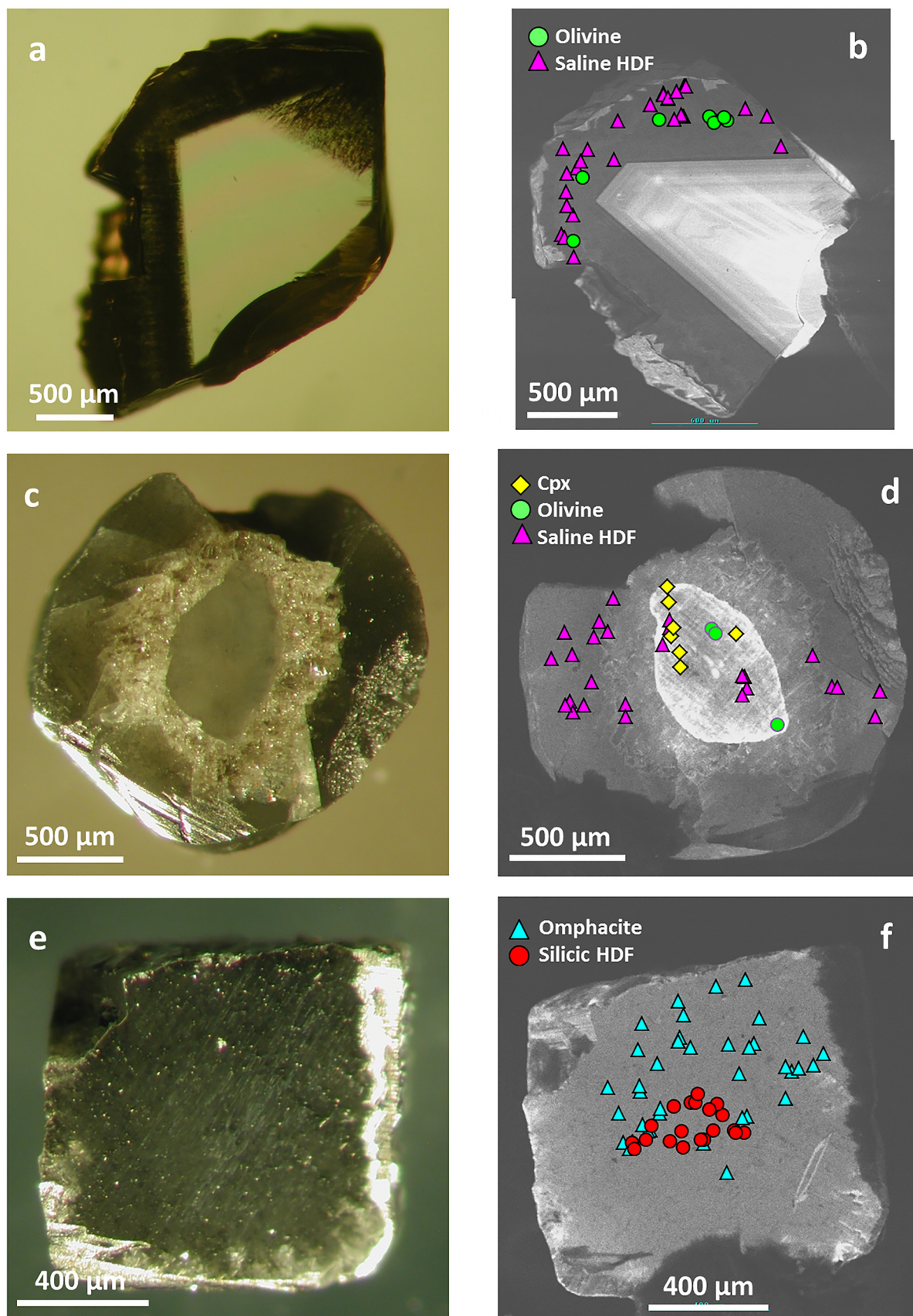
applicable from one batch of samples to another. Any analyte below the LOQ is flagged in the data and not used on a concentration plot. In the previous definition³⁹, data can only be quantitative if it exceeds 10σ of the blank, hence the analyte/blank ratio is a critical parameter to measure. The total amount of analyte and hence the analyte/background ratio is simply a function of the length of the ablation, with the ratio increasing with time.

Multi-element ICP-MS. TPBs and aliquot sample solutions were analysed for trace element concentrations on the Thermo-Electron Element II ICPMS at Durham University. Each sample aliquot was made to 500 µl with 3% HNO₃. Instrumental conditions were similar to those described previously⁴⁰. Solution concentrations were measured against 9-point calibration lines constructed from appropriately dilute solutions of the international standards AGV-1, BHVO-1 and W-2. All concentrations were corrected for instrument drift using an 115In internal spike. Oxide correction coefficients were determined by running standard solutions of Ba, La, Ce, Pr, Nd, Sm, Gd and Tb at the beginning of each analytical session to correct for the daily changes in the oxide production rate. All trace element concentrations were normalized to the diamond weight loss during ablation.

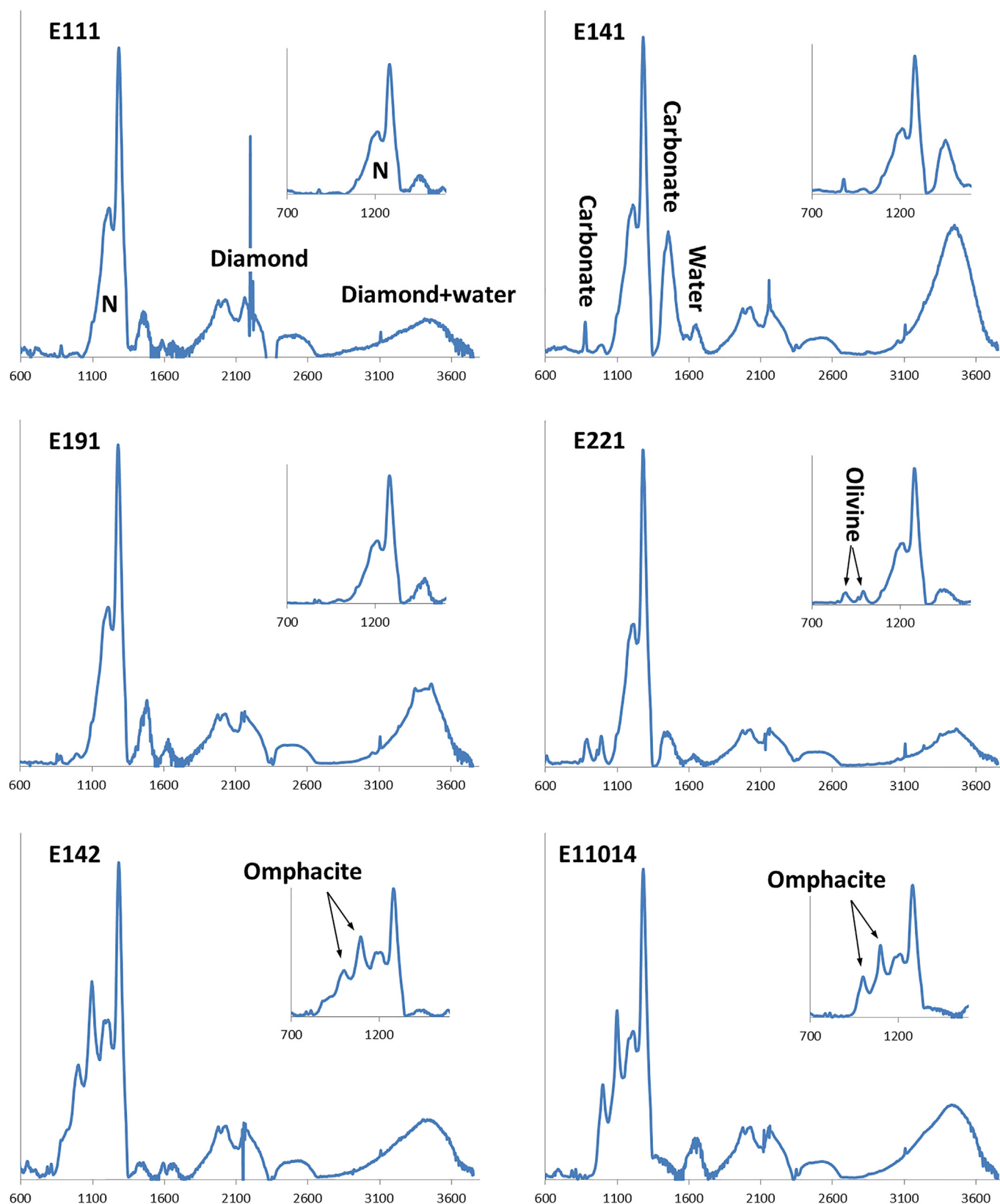
Thermal ionization mass spectrometry. With each batch of samples processed for isotopic analysis, between five and ten TPBs were carried out to determine the average size of the blank contribution and its effect on the isotopic composition of the sample. During the course of this study Sr blanks averaged 5 pg (n = 12). A Sr isotope blank correction was performed using a measured blank isotopic composition based on combining the equivalent of over 60 TPBs to yield sufficient Sr (~500 pg) for a precise and accurate thermal ionization mass spectrometry (TIMS) analysis. The average ⁸⁷Sr/⁸⁶Sr composition of the laboratory blank during the course of this work was 0.710853 ± 0.000194 and all Sr samples were blank-corrected based on this value and the average blank set at 5 pg. Sr samples were loaded using procedures described in detail previously^{37,40}, employing a purified TaF₅ activator. Sr isotope ratios were measured on a ThermoFisher Triton TIMS at Durham University. Sr isotope measurements were carried out using a static multi-collection routine. Each sample measurement achieved between 50 and 300 ratios with an integration time of 4 s per ratio; total analysis time was approximately 3–20 min. Mass fractionation was corrected using an exponential law and an ⁸⁶Sr/⁸⁸Sr ratio of 0.1194. Multiple loads (n = 43) of NBS987 of between 0.5 and 3 ng size gave an average value of 0.710260 ± 0.00002 (2 standard deviations; n = 43), which compares well to the long-term values reported from the Durham University laboratory for similar sized standards from the same laboratory^{37,38,40,41}. As the Durham laboratory reports Sr data relative to an ⁸⁷Sr/⁸⁶Sr ratio of 0.710240 no additional normalization was performed. Average signal size for the ⁸⁸Sr for the 0.5 ng and 3 ng standards were 0.8 ± 0.4 V and 5 ± 1.3 V, respectively. Signal sizes for samples were on average 0.2 ± 1 V. We have previously documented in detail the levels of accuracy and repeatability for samples and standards at these low signal intensities³⁸. There is no systematic relationship between analyte size and Sr isotope composition after blank correction. Hence we conclude that our blank correction procedures adequately correct for our systematic TPB. Uncertainties in the magnitude and isotopic composition of the blank are incorporated into the reported errors on isotopic compositions at the 2σ level. Previous experiments³⁸ indicate that for blanks of ~5 pg, it is possible to make accurate blank corrections to samples containing as little as 20 pg and therefore that level was used as a cut-off for accepting accurate data in this study because similar levels of blank reproduction were achieved.

- Boyd, S. R., Kiflawi, I. & Woods, G. S. The relationship between infrared absorption and the A defect concentration in diamond. *Philos. Mag. B* **69**, 1149–1153 (1994).
- Boyd, S. R., Kiflawi, I. & Woods, G. S. Infrared absorption by the B nitrogen aggregation in diamond. *Philos. Mag. B* **72**, 351–361 (1995).
- Kiflawi, I., Myers, A. E., Spear, P. M., van Wyk, J. A. & Woods, G. S. Infrared absorption by the single nitrogen and A defect centres in diamond. *Philos. Mag. B* **69**, 1141–1147 (1994).
- Zaitsev, A. M. *Optical Properties of Diamond Handbook* 502 (Springer, 2001).
- Weiss, Y., Kiflawi, I. & Navon, O. IR spectroscopy: quantitative determination of the mineralogy and bulk composition of fluid microinclusions in diamonds. *Chem. Geol.* **275**, 26–34 (2010).
- Bastin, G. & Heijligers, J. in *Electron Probe Quantitation, Workshop at the National Bureau of Standards, Gaithersburg, Maryland* (eds Heinrich, K. & Newbury, D.) 145–161 (Plenum, 1991).
- Charlier, B. L. A. et al. Methods for the microsampling and highprecision analysis of strontium and rubidium isotopes at single crystal scale for petrological and geochronological applications. *Chem. Geol.* **232**, 114–133 (2006).
- Harlou, R., Pearson, D. G., Nowell, G. M., Ottley, C. J. & Davidson, J. P. Combined Sr isotope and trace element analysis of melt inclusions at sub-ng levels using micro-milling, TIMS and ICPMS. *Chem. Geol.* **260**, 254–268 (2009).
- Currie, L. A. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal. Chem.* **40**, 586 (1968).

40. Font, L., Nowell, G. N., Pearson, D. G., Ottley, C. J. & Willis, S. G. Sr isotope analysis of bird feathers by TIMS: a tool to trace bird migration paths and breeding sites. *J. Anal. At. Spectrom.* **22**, 513–522 (2007).
41. Klein-BenDavid, O. *et al.* Mixed fluid sources involved in diamond growth constrained by Sr-Nd-Pb-C-N isotopes and trace elements. *Earth Planet. Sci. Lett.* **289**, 123–133 (2010).
42. Stachel, T. & Harris, J. W. The origin of cratonic diamonds—constraints from mineral inclusions. *Ore Geol. Rev.* **34**, 5–32 (2008).
43. Taylor, W. R., Canil, D. & Milledge, J. Kinetics of Ib to IaA nitrogen aggregation in diamond. *Geochim. Cosmochim. Acta* **60**, 4725–4733 (1996).
44. Taylor, W. R., Jaques, A. L. & Ridd, M. Nitrogen-defect aggregation characteristics of some Australasian diamonds: time–temperature constraints on the source regions of pipe and alluvial diamonds. *Am. Mineral.* **75**, 1290–1310 (1990).
45. Klein-BenDavid, O., Izraeli, E. S., Hauri, E. & Navon, O. Mantle fluid evolution—a tale of one diamond. *Lithos* **77**, 243–253 (2004).
46. Sarkar, C., Heaman, L. M. & Pearson, D. G. Duration and periodicity of kimberlite volcanic activity in the Lac de Gras kimberlite field, Canada and some recommendations for kimberlite geochronology. *Lithos* **218–219**, 155–166 (2015).

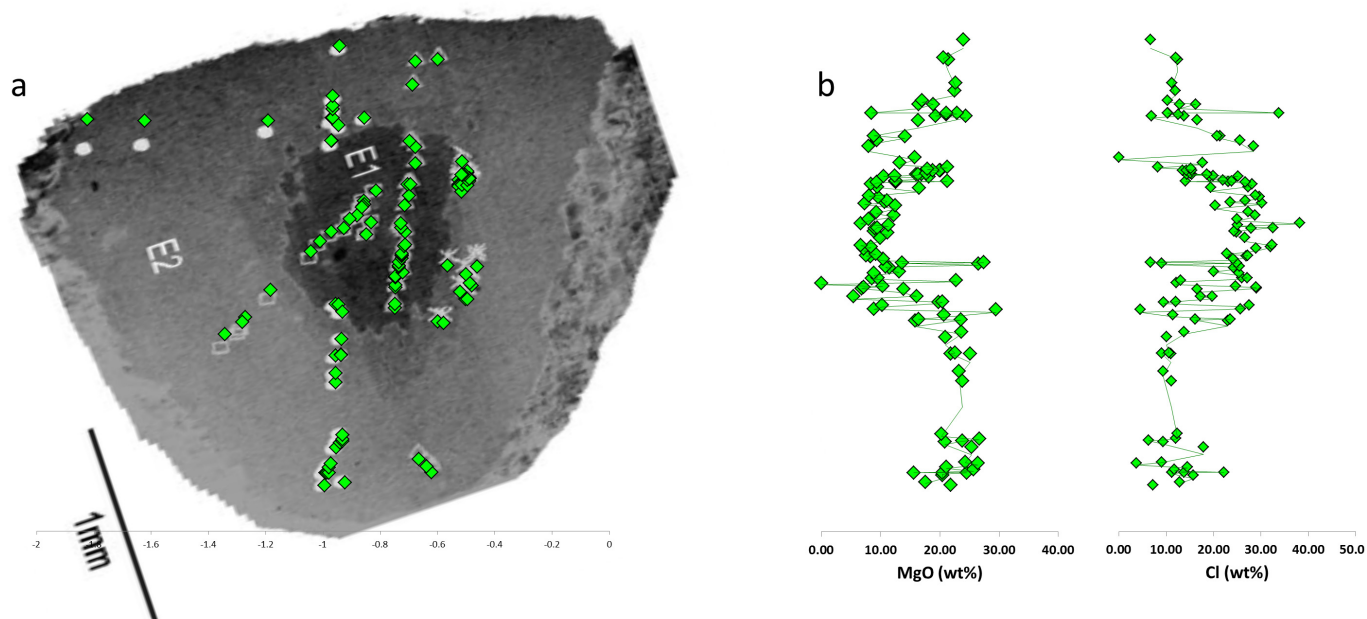


Extended Data Figure 1 | Photomicrographs and cathodoluminescence images of three diamonds from the Fox kimberlite. a, b, Diamond E154; c, d, diamond E217; e, f, diamond E142. The location of the microinclusions analysed by EPMA is imposed on the cathodoluminescence images.



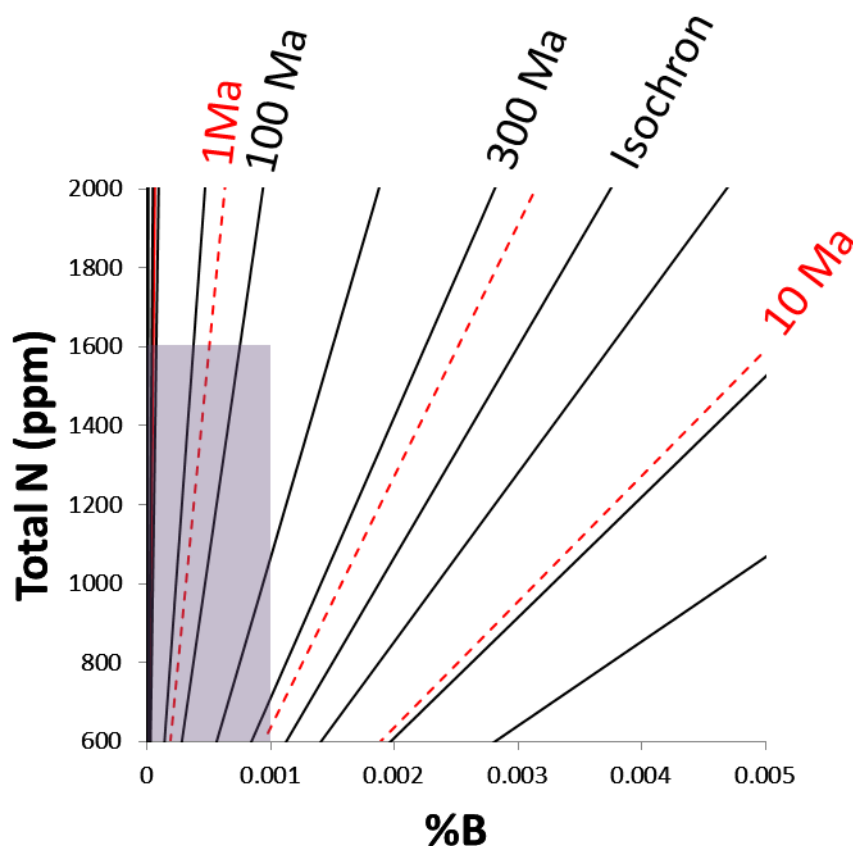
Extended Data Figure 2 | Infrared spectra of six fibrous diamonds from the Fox kimberlite. Absorbance bands of the diamond itself, nitrogen in the diamond lattice, and carbonate and water in the microinclusions are observed in all diamonds. In diamond E221 olivine absorption due to the presence of olivine microinclusions is also observed, while in diamonds E142 and E11014 the presence of pyroxene (omphacite) is detected. The nitrogen in all Fox fibrous diamonds is aggregated in 100% A centres. The nitrogen defect structure in diamonds evolves over time, from individual nitrogen atoms

replacing single carbon atoms (C centres) when the diamond forms, to pairs of adjacent nitrogen atoms (A centres), to four nitrogen atoms tetrahedrally arranged around a vacancy (B centres). At the mantle temperatures at which most diamonds form (that is, 950–1,250 °C)⁴², the full transformation from C to A centres (100% A centres) will form over a timescale of hundreds of thousands to millions of years, while A centres merge to form B centres over billions of years^{43,44}.



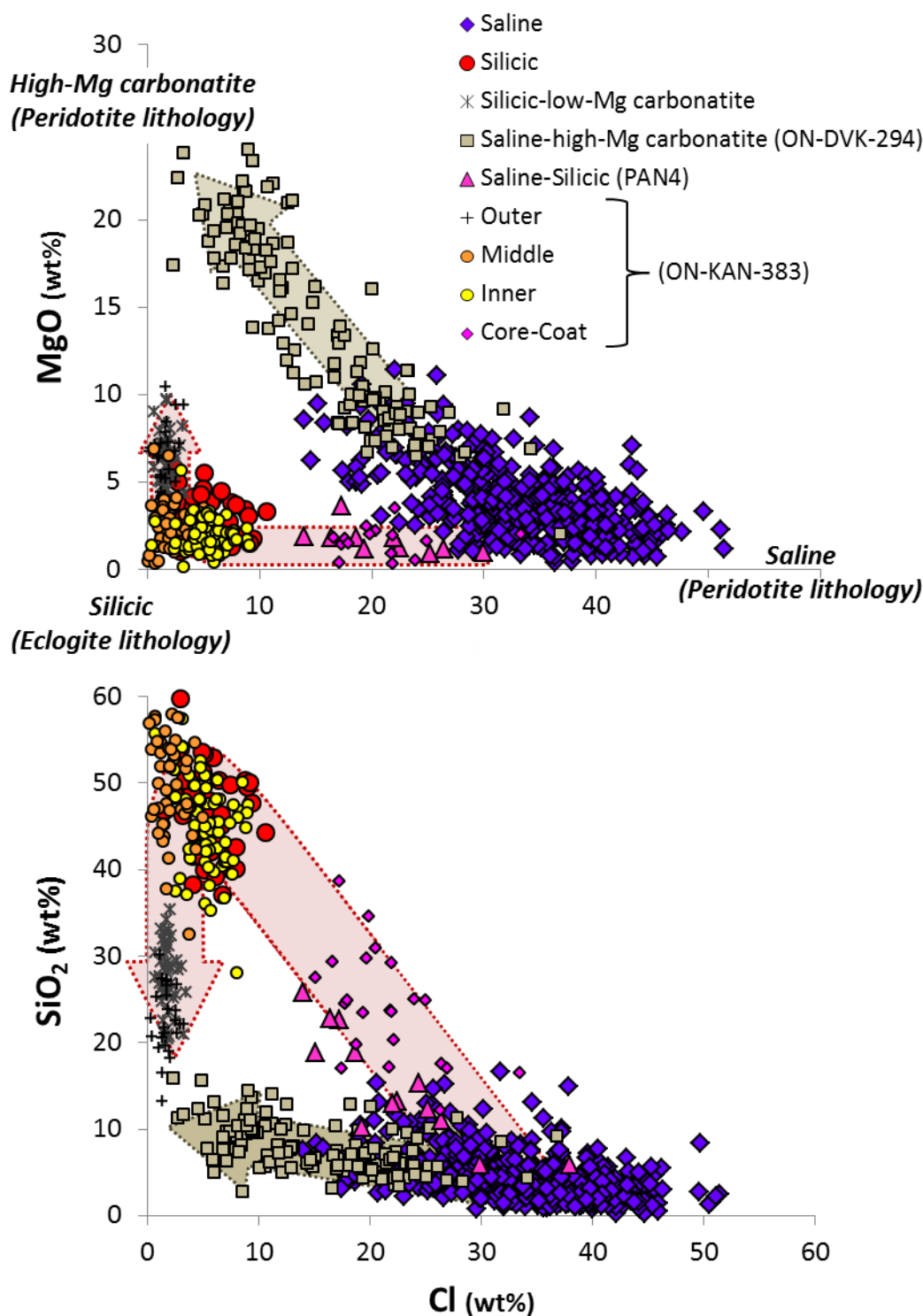
Extended Data Figure 3 | Microinclusion spatial location and MgO and Cl composition in diamond ON-DVK-294 (refs 7, 45). **a**, The data point locations are superimposed on the original figure 3 from ref. 45 (reproduced with permission). The shift in the position of a few of the data points is related to errors in the position of fiducial marks. **b**, The position of each data

point was radially normalized to allow plotting of the compositional change of the fluids along a uniform diamond profile. Clear chemical differences between the inner and outer parts of this diamond exist; however, the transition between the two zones in the diamond shows a gradual evolution rather than an abrupt change.



Extended Data Figure 4 | Total N in parts per million (ppm) versus percentage of B centres for fibrous diamonds from the central Slave lithosphere. Isochrons are calculated for ambient temperatures of 930 °C (dotted red lines) and 1,010 °C (solid black lines) as described previously^{43,44}. Fibrous diamonds in the Slave lithosphere have 600–1,600 ppm nitrogen and are essentially of pure IaA type (that is, %B = 0). The presence of a very small quantity of B centres ($\ll 1$) is difficult to evaluate from FTIR spectroscopy but

such centres are necessary for calculating a possible age based on nitrogen aggregation. The geothermometry results for mineral inclusions in fibrous diamonds from the Panda kimberlite yielded low equilibrium temperatures of 930–1,010 °C (ref. 5). Applying these temperatures to the amount of nitrogen in the central Slave diamonds and %B ≈ 0 (grey area), as indicated from FTIR spectroscopy, reasonable mantle residence times for these diamonds vary between 1–210 Ma, before the kimberlite eruption age of 55 Ma (ref. 46).



Extended Data Figure 5 | MgO and SiO₂ against Cl content, all in wt%, water- and CO₂-free composition, of HDF microinclusions in fibrous diamonds from the central Slave craton and diamond ON-KAN-383 from Kankan, Guinea. Intermediate compositions between saline to silicic are not unique to the PAN4 diamond⁵. The composition of 21 microinclusions with similar composition in an eclogitic zoned coated diamond from Kankan, Guinea (ON-KAN-383) was previously reported⁷. The saline-to-silicic inclusions in diamond ON-KAN-383 are restricted to the core-coat boundary of this diamond and develop, progressing towards silicic compositions

(inner, middle) that finally evolve to low-Mg carbonatitic HDFs close to the rim of the diamond (outer). The paragenesis of this diamond and its analogous HDF compositions to eclogitic diamonds from the central Slave craton provide strong evidence for broader geographical involvement of saline fluids in the formation of silicic melts in eclogite, as well as for the continuous evolution (shaded red arrow) from saline through silicic to low-Mg carbonatitic composition when saline HDFs intersect with hydrous-carbonated eclogite lithology in the CLM root.

Reconstructing the reproductive mode of an Ediacaran macro-organism

Emily G. Mitchell¹, Charlotte G. Kenchington^{1,2}, Alexander G. Liu³, Jack J. Matthews⁴ & Nicholas J. Butterfield¹

Enigmatic macrofossils of late Ediacaran age (580–541 million years ago) provide the oldest known record of diverse complex organisms on Earth, lying between the microbially dominated ecosystems of the Proterozoic and the Cambrian emergence of the modern biosphere¹. Among the oldest and most enigmatic of these macrofossils are the Rangeomorpha, a group characterized by modular, self-similar branching and a sessile benthic habit^{2–4}. Localized occurrences of large *in situ* fossilized rangeomorph populations allow fundamental aspects of their biology to be resolved using spatial point process techniques⁵. Here we use such techniques to identify recurrent clustering patterns in the rangeomorph *Fractofusus*, revealing a complex life history of multigenerational, stolon-like asexual reproduction, interspersed with dispersal by waterborne propagules. Ecologically, such a habit would have allowed both for the rapid colonization of a localized area and for transport to new, previously uncolonized areas. The capacity of *Fractofusus* to derive adult morphology by two distinct reproductive modes documents the sophistication of its underlying developmental biology.

Late Ediacaran sedimentary strata (~580–541 million years ago (Ma)) of Newfoundland and the UK are dominated by rangeomorphs, whose unique self-similar branching construction³ makes resolution of their phylogenetic relationships, or even their basic biology, difficult¹. The occurrence of rangeomorphs in conspicuously deep-water sediments has led to a consensus that they were heterotrophic⁶, while the global distribution of charniids (a rangeomorph sub-group) has been interpreted as evidence for reproduction via waterborne propagules⁷. In the present study we use spatial statistics and modelling^{5,8} in a novel approach to illuminate the reproductive biology and underlying ecology of one of the most abundantly preserved rangeomorph fossils, *Fractofusus*⁹.

We analysed three large bedding-plane assemblages of *Fractofusus* in southeast Newfoundland: (1) the 'D' surface and (2) the 'E' surface at Mistaken Point, Avalon Peninsula^{9,10} and (3) the H14 surface on Bonavista Peninsula (locality 14 of ref. 11) (Extended Data Fig. 1a–c). A volcanic tuff directly above the 'E' surface has been dated to 565 ± 3 Ma (ref. 12), which also constrains the age of the underlying 'D' surface. Regional lithostratigraphic correlations suggest that the H14 surface is a few million years younger than the Mistaken Point beds¹¹. All three assemblages occur within deep-marine turbidite sequences, with *Fractofusus* fossils preserved as negative epirelief external moulds in siltstone hemipelagites, cast from above by volcanoclastic deposits⁶.

Fractofusus is conspicuously endemic, restricted almost exclusively to southeastern Newfoundland¹³, where it dominates many macrofossil assemblages¹⁰. *Fractofusus* has a rounded, elongate spindle-like morphology, with two (arguably three^{2,13}) offset rows of irregularly alternating, self-similar, subdivided frondlets arranged along a central axis^{2,14}. *Fractofusus* specimens range from 1 cm to 42 cm in length² (Fig. 1a, b); two species have been described, distinguished by their length:width (L/W) ratios². The 'D' and 'E' surfaces are dominated by

the elongate form, *Fractofusus misrai* (L/W = 3.2; Fig. 1b), whereas the more ovate *Fractofusus andersoni* (L/W = 1.6; Fig. 1a) dominates the H14 surface¹⁰. *Fractofusus* occurs in dense benthic populations and exhibits no evidence of motility or current orientation². Together with nearest neighbour spatial analyses¹⁰, these observations point to a sessile, recumbent, benthic mode of life in aggregated communities.

The spatial positions of *Fractofusus* were mapped to millimetre-scale resolution using differentiated Global Positioning System (GPS) (Extended Data Fig. 1d, e) on the two surfaces at Mistaken Point, and by tracing specimen outlines onto acetate sheets at H14; importantly,

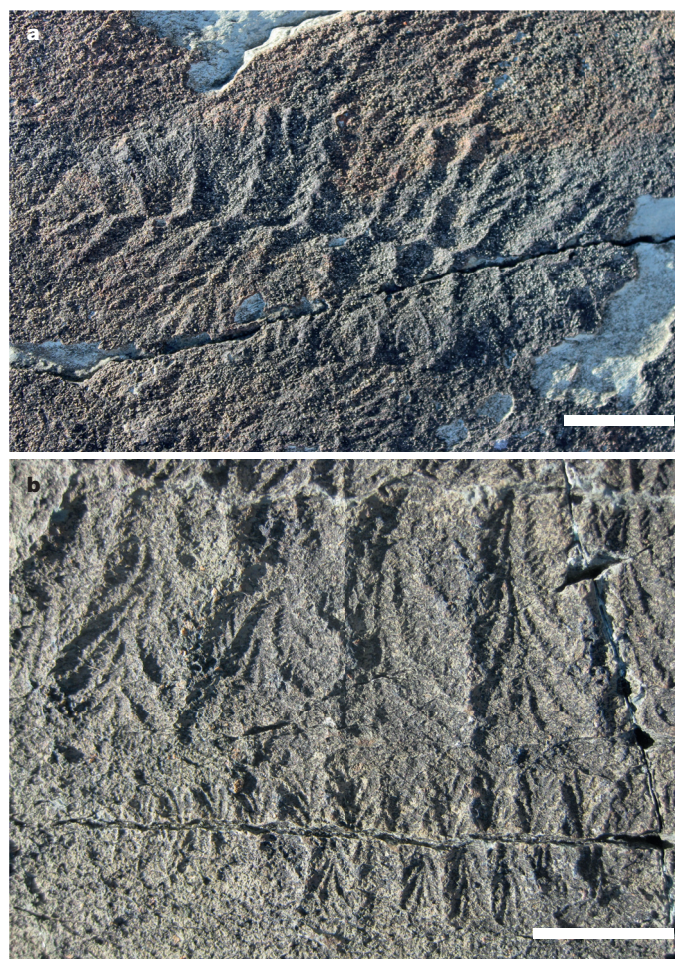


Figure 1 | *Fractofusus* specimens from Newfoundland, Canada. **a**, *F. andersoni* specimen from the H14 surface. **b**, *F. misrai* from the 'E' surface, showing a large size-class partial specimen (~20 cm, above) alongside a small size-class specimen (3.5 cm in length, below). Scale bars, 1 cm. Photographs are non-retrodeformed.

¹Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK. ²British Geological Survey, Keyworth, Nottingham NG12 5GG, UK. ³School of Earth Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS8 1TQ, UK. ⁴Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3AN, UK.

this latter approach also allowed size data to be recorded (Extended Data Fig. 1f). The 'D' and 'E' surface data were corrected for tectonic deformation before analysis (Extended Data Fig. 2)⁷. Heterogeneous Poisson models were used to identify possible distortions arising from

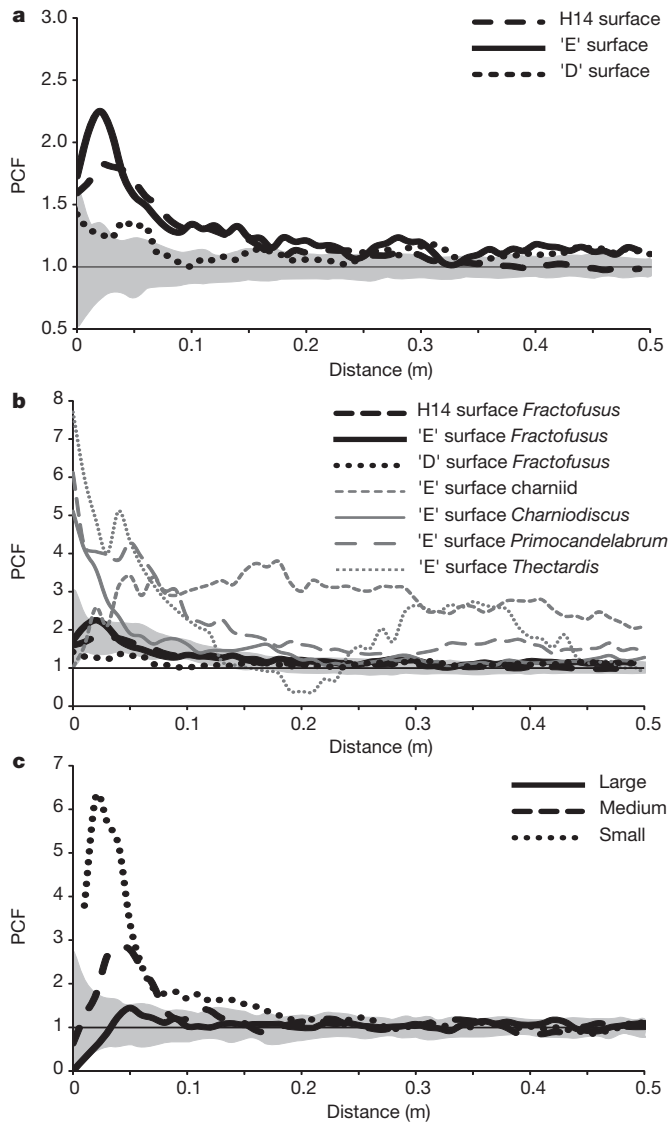


Figure 2 | PCF for mapped taxa. For all plots the x axis is the inter-point distance between organisms in metres. On the y axis, PCF = 1 indicates CSR, <1 indicates segregation and >1 indicates aggregation. **a**, PCF for *Fractofusus* on the 'D' surface (1,040 specimens), 'E' surface (1,141 specimens) and H14 surface (1,214 specimens). Grey shaded area depicts the bounds of 99 Monte Carlo simulations of CSR. Since the PCF curves are not completely within these areas, the CSR hypothesis is rejected and one can assume that the *Fractofusus* distributions on all three surfaces form cluster patterns ($p_d^D < 0.01$, $p_d^E < 0.01$, $p_d^{H14} < 0.01$). **b**, PCF for non-CSR 'E' surface taxa (charniid 76 specimens, *Charniodiscus* 326 specimens, *Primocandelabrum* 311 specimens and *Thectardis* 39 specimens). Grey shaded area depicts 99 Monte Carlo simulations of the best-fit H14 surface model of double Thomas cluster process. Note how the 'E' surface *Fractofusus* PCF follows the H14 surface PCF very closely, and can be modelled by the same process ($p_d = 0.51$). Other 'E' surface taxa have markedly different PCF to the *Fractofusus* PCF. **c**, PCF for the three size-classes of *Fractofusus* on H14 surface. Grey shaded area depicts the 99 Monte Carlo simulations of CSR. The large size-class (350 specimens) exhibits CSR ($p_d = 0.30$); the intermediate size-class (310 specimens) shows aggregation <0.10 m (single Thomas cluster model, $p_d = 0.51$). The small size-class (554 specimens) shows a large aggregation <0.08 m and a lesser aggregation between 0.08 m and 0.20 m (double Thomas cluster model, $p_d = 0.72$).

differential erosion of the bedding planes (Supplementary Table 1). Pair correlation functions (PCFs) were calculated to describe the spatial distributions of taxa on each bedding plane (Fig. 2a)⁵. Monte Carlo simulations¹⁵ and Diggle's goodness-of-fit test² (the p -value p_d , in which $p_d = 1$ indicates a perfect model fit, and $p_d = 0$ indicates no fit), were used to compare the fit of different spatial models to the data (specifically homogeneous and heterogeneous Poisson models¹⁶ and single and double homogeneous and heterogeneous Thomas cluster models¹⁶). PCFs were also used to describe the spatial distributions of taxa other than *Fractofusus* on the 'D' and 'E' surfaces (Fig. 2b). For the H14 surface, spatial relationships between three distinct *Fractofusus* size-classes (defined in Methods; Extended Data Fig. 3a, b) were analysed by calculating partial PCFs⁵ and comparing model fit of bivariate shared parent (SP) models with linked cluster models (LCMs) (Fig. 2c)¹⁶. Finally, spatial directionality was investigated by plotting the generalized K -functions¹⁷ of *Fractofusus* specimens from H14 surface from 0° to 360° (isotropy plots), allowing visualization of the relative directional positions of specimens (Fig. 3).

Non-random spatial distributions of sessile organisms, namely those that do not exhibit complete spatial randomness (CSR), can be explained by either extrinsic factors (for example, environmental heterogeneities), or intrinsic reproduction¹⁸. Identifying the processes behind such patterns is not straightforward; however, extrinsically induced patterns are generally best modelled by heterogeneous Poisson models¹⁸, which describe randomly distributed points with a non-uniform density across the sampled area. In contrast, intrinsic processes typically generate Thomas cluster models¹⁸, where the points within each cluster have a normal density distribution centred on a parent point.

All three populations of *Fractofusus* were found to be significantly aggregated, conforming closely to homogeneous Thomas cluster models (Fig. 2a). Specimens on the 'E' and H14 surfaces are aggregated at two spatial scales, forming clusters of clusters (Fig. 2a, b). On the 'E' surface, this distribution is best modelled by a nested homogeneous double Thomas cluster model of 23 clusters (radius $r = 0.242$ m), each containing 12 smaller clusters ($r = 0.074$ m) of 3 specimens ($p_d = 0.76$). The H14 surface distribution is best modelled by a nested homogeneous double Thomas cluster model of 24 large clusters ($r = 0.237$ m), each containing 6 clusters ($r = 0.079$ m) of 8 specimens ($p_d = 0.89$). The 'D' surface distribution forms discrete clusters (not clusters of clusters), which are best modelled by a single Thomas cluster model ($p_d = 0.77$) with 338 *Fractofusus* clusters of 3 specimens ($r = 0.086$ m) (Extended Data Tables 1 and 2). Importantly, the spatial distribution on the 'E' surface can also be modelled by the same nested double cluster pattern as found on the H14 surface (Fig. 2b) ($p_d^{H \text{ on } E} = 0.51$), strongly implying the same underlying process for both distributions (Fig. 2b and Extended Data Table 3). The spatial distribution of *Fractofusus* on the 'D' surface is conspicuously similar to that seen in the larger specimens on H14 (Extended Data Fig. 4e). By contrast, the spatial distributions of other taxa—*Thectardis*, *Primocandelabrum* and *Charniodiscus*—exhibit fundamentally different magnitudes and spatial scales of aggregation, both to each other and to those of *Fractofusus* (Fig. 2b and Extended Data Tables 4 and 5).

The close fit of *Fractofusus* spatial distributions to single and nested double Thomas cluster models strongly suggests that they derive from reproductive rather than extrinsic (environmental) factors. Reproductive biology is further corroborated by size analysis of the *Fractofusus* population on the H14 surface (Fig. 2c and Extended Data Fig. 4a, c–e), which reveals strikingly different spatial patterns for each of the three size-classes (Fig. 2c, Extended Data Tables 1 and 2). Whereas the largest size-class (>11.0 cm) is randomly distributed ($p_d = 0.30$), both the intermediate (5.5–11.0 cm) and smallest (<5.5 cm) size-classes are hierarchically clustered: small individuals cluster around intermediate individuals ($p_d^{LCM} = 0.74$ versus $p_d^{SP} = 0.03$; Extended Data Table 5), which in turn cluster around large individuals ($p_d^{LCM} = 0.66$ versus $p_d^{SP} = 0.01$). In other words,

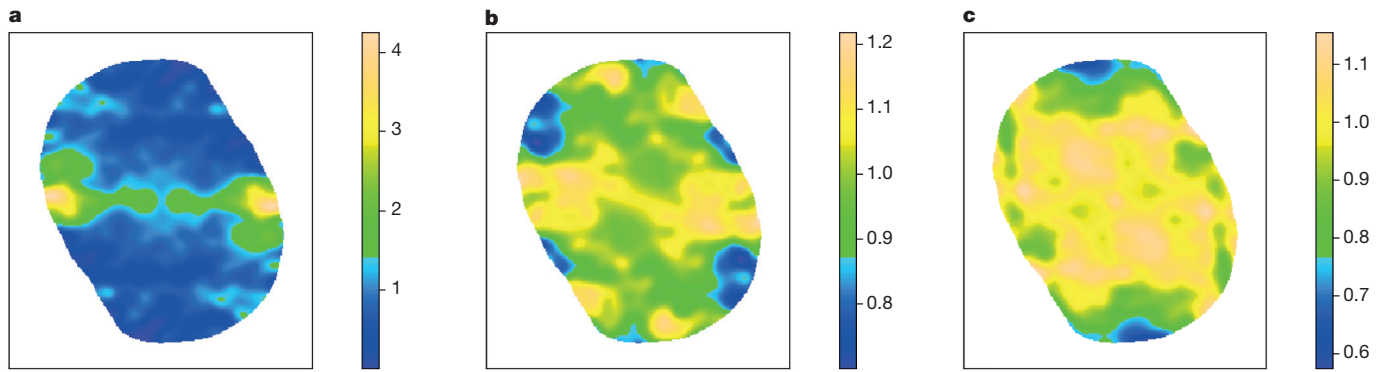


Figure 3 | Isotropy plots from the H14 surface for each size-class of *Fractofusus*, providing a visualization of specimen positions relative to one another. The vertical axis on each part depicts the colour map of specimens per square metre normalized to account for different densities between size-classes. A peak (>1) is shown in green or yellow and depicts aggregation, while a dip (<1) is shown in blue and depicts segregation. If there are no directional

effects then the colour map in every direction from the centre point should be similar. **a**, The large size-class shows strong anisotropy, with aggregation of up to four normalized specimens per square metre. **b**, **c**, In contrast the medium (**b**) and small (**c**) size-classes show isotropy, namely a relative evenness of aggregations with a maximum density variation up to 0.5 normalized specimens per square metre.

the smallest specimens form clusters (homogeneous nested double Thomas cluster model, $p_d = 0.72$) around intermediate-sized specimens (homogeneous single Thomas cluster model, $p_d = 0.51$), which are themselves clustered around randomly distributed large specimens (homogeneous Poisson model, $p_d = 0.31$; Figs 2c and 4, Extended Data Fig. 5 and Extended Data Tables 1, 2 and 5). Moreover, the isotropy plots for H14 (Fig. 3) show strong directionality for the large size-class, but limited directionality for the medium and small size-classes. The nested clusters on the 'E' and H14 surfaces suggest three generations, while the single clusters on the 'D' surface suggest two, reflecting an earlier stage in population development. The difference is consistent with the suggestion that the 'D' surface records an earlier stage in the ecological succession of Ediacaran macroscopic communities¹⁰ (Supplementary Information section 2.3).

As with other Ediacaran macrofossils, there is no direct fossil evidence of reproductive habits in *Fractofusus*, but its recurrent distribution on bedding surfaces provides a statistically robust approach for inferring the underlying processes¹⁹. In modern oceans, large sessile organisms typically reproduce by means of waterborne propagules, fragmentation/budding and/or stolons (that is, production of asexual

clones that are at least initially connected to the parent by specialized outgrowths).

Spatial distribution of waterborne propagules—including both sexual and asexual spores, as well as sub-millimetre buds and fragments—is a function of current and rate of sinking. Even with rapid sinking ($\sim 1 \text{ mm s}^{-1}$)²⁰ and slow currents ($\sim 1 \text{ cm s}^{-1}$), propagules released from the dorsal surface of a 'parental' *Fractofusus* ($\sim 2\text{--}3 \text{ cm}$ above the substrate) would have been current aligned²¹ and dispersed by decimetres or more²². Slow descent times also correspond with right-skewed (mean greater than the median) density distributions^{20,21}. The random spatial distribution of the H14 largest size-class most probably reflects a large dispersal distance (Extended Data Table 1), which, coupled with its highly directional isotropy plot (Fig. 3a), indicates that the largest specimens were strongly influenced by currents (cf. refs 7, 10). As such, they probably derive from waterborne propagules and represent the initial establishment of a *Fractofusus* population on this surface.

The hierarchically clustered bedding plane distributions of small and medium *Fractofusus* on H14 closely match patterns exhibited by organisms reproducing asexually via stolon-like lateral extensions (Extended Data Fig. 5)²³. Cluster distributions of the small and medium size-classes are also highly left-skewed (median greater than the mean), with the mean distance from each 'parent' to their 'offspring' on the order of a few centimetres (Fig. 2a), and offspring exhibiting no significant directionality or current orientation (Fig. 3b–c, Extended Data Fig. 3c). The reproducibility of the model distributions across the three bedding-plane assemblages further attests to the indifferent effects of current: the spatial distributions of non-tethered offspring would result in patterns dependent on current velocity, which are unlikely to be consistent across multiple bedding planes in different localities. Moreover, there are no recorded instances of buds or fragmentary specimens of *Fractofusus* in any of its 5,000+ documented specimens^{1,10,11,24} (see Supplementary Information section 2.1 and 2.5). As such, the *Fractofusus* clusters on the H14 surface are not consistent with waterborne propagules or fragmentation/budding, but are directly comparable to stolon-like reproduction. Other taxa exhibit an intriguing range of non-random habits, and our preliminary analyses indicate that *Primocandelabrum* and *Charniodiscus* may have also reproduced using stolons.

Reproductive biology lies at the core of ecological and evolutionary dynamics, and its positive identification in Ediacaran macrofossils has the potential to illuminate the beginnings of the modern marine biosphere. Previous studies of Ediacaran macrofossils have investigated the seasonality of reproduction⁷, identified putative stolons^{25,26} and inferred sexual or asexual reproduction on the basis of biogeographic distribution or qualitative description of local populations^{7,10,27–29}. In

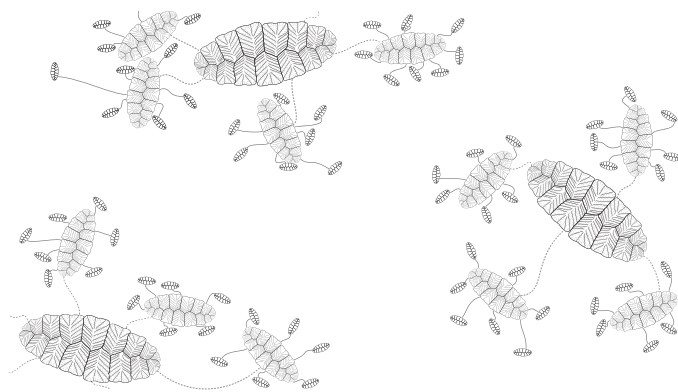


Figure 4 | Schematic diagram showing simplified *Fractofusus* spatial arrangements. The actual number of clusters, and clusters within those clusters, is higher than shown (23 clusters each containing 12 clusters of 3 specimens on the H14 surface), making their direct visual detection challenging. No overlapping specimens are shown because, while the best-fit models allow for overlaps, the observed PCF between the small size-class (Extended Data Fig. 4c, d) and the large size-class (Fig. 2b) shows a small segregation ($<3 \text{ cm}$) away from the model behaviour, and a similar, non-significant segregation for the large size-class.

the case of phosphatized 'embryo' microfossils, internal cell packages have been interpreted as evidence of germ–soma differentiation³⁰, but it remains to be seen how those fossils relate to the evolution of large and/or complex eukaryotes. The identification in *Fractofusus* of a multigenerational asexual clonal phase, interspersed with the release of waterborne propagules, is the first statistically robust account of reproductive life history reported in an Ediacaran macrofossil. Such a strategy would have allowed for the rapid exploitation of localized areas, as well as for transport to new, previously uncolonized areas. The conclusion that *Fractofusus* could switch between reproductive modes further reveals the sophistication of its underlying developmental programme, capable not only of tissue differentiation, but also the generation of new macroscopic individuals both from benthic stolons and from waterborne propagules.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 February; accepted 8 June 2015.

Published online 3 August 2015.

- Liu, A. G., Kenchington, C. G. & Mitchell, E. G. Remarkable insights into the paleoecology of the Avalonian Ediacaran biota. *Gondwana Res.* **27**, 1355–1380 (2015).
- Gehling, J. G. & Narbonne, G. M. Spindle-shaped Ediacara fossils from the Mistaken Point assemblage, Avalon Zone, Newfoundland. *Can. J. Earth Sci.* **44**, 367–387 (2007).
- Narbonne, G. M. Modular construction of early Ediacaran complex life forms. *Science* **305**, 1141–1144 (2004).
- Hoyal Cuthill, J. F. & Conway Morris, S. Fractal branching organizations of Ediacaran rangeomorph fronds reveal a lost Proterozoic body plan. *Proc. Natl Acad. Sci. USA* **111**, 13122–13126 (2014).
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. *Statistical Analysis and Modelling of Spatial Point Patterns* Vol. 70 (John Wiley, 2008).
- Wood, D. A., Dalrymple, R. W., Narbonne, G. M., Gehling, J. G. & Clapham, M. E. Paleoenvironmental analysis of the late Neoproterozoic Mistaken Point and Trepassy formations, southeastern Newfoundland. *Can. J. Earth Sci.* **40**, 1375–1391 (2003).
- Darroch, S. A. F., Laflamme, M. & Clapham, M. E. Population structure of the oldest known macroscopic communities from Mistaken Point, Newfoundland. *Paleobiology* **39**, 591–608 (2013).
- Wiegand, T., Gunatilleke, S., Gunatilleke, N. & Okuda, T. Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology* **88**, 3088–3102 (2007).
- Landing, E., Narbonne, G. M. & Myrow, P. (eds) Trace fossils, small shelly fossils and the Precambrian–Cambrian boundary. *Bull. NY State Mus.* **463**, 1–81 (1988).
- Clapham, M. E., Narbonne, G. M. & Gehling, J. G. Paleoeology of the oldest known animal communities: Ediacaran assemblages at Mistaken Point, Newfoundland. *Paleobiology* **29**, 527–544 (2003).
- Hofmann, H. J., O'Brien, S. J. & King, A. F. Ediacaran biota on Bonavista Peninsula, Newfoundland. *Can. J. Paleontol.* **82**, 1–36 (2008).
- Benus, A. P. Sedimentological context of a deep-water Ediacaran fauna (Mistaken Point Formation, Avalon zone, eastern Newfoundland). *Bull. NY State Mus.* **463**, 8–9 (1988).
- Narbonne, G. M., Laflamme, M., Trusler, P. W., Dalrymple, R. W. & Greentree, C. Deep-water Ediacaran fossils from northwestern Canada: Taphonomy, ecology, and evolution. *J. Paleontol.* **88**, 207–223 (2014).
- Brasier, M. D., Antcliffe, J. B. & Liu, A. G. The architecture of Ediacaran fronds. *Palaeontology* **55**, 1105–1124 (2012).
- Diggle, P. *Statistical Analysis of Spatial Point Patterns* 2nd edn (Arnold, 2003).
- Baddeley, A. & Turner, R. Practical maximum pseudolikelihood for spatial point patterns. *Aust. NZ J. Stat.* **42**, 283–322 (2000).
- Chiu, S. N., Stoyan, D., Kendall, W. S. & Mecke, J. *Stochastic Geometry and its Applications* 3rd edn (John Wiley, 2013).
- Lin, Y., Chang, L., Yang, K., Wang, H. & Sun, I. Point patterns of tree distribution determined by habitat heterogeneity and dispersal limitation. *Oecologia* **165**, 175–184 (2011).
- Droser, M. L. & Gehling, J. G. Synchronous aggregate growth in an abundant new Ediacaran tubular organism. *Science* **319**, 1660–1662 (2008).
- Gaylord, B., Reed, D. C., Raimondi, P. T. & Washburn, L. Macroalgal spore dispersal in coastal environments: mechanistic insights revealed by theory and experiment. *Ecol. Monogr.* **76**, 481–502 (2006).
- Shanks, A. L. Pelagic larval duration and dispersal distance revisited. *Biol. Bull.* **216**, 373–385 (2009).
- Gaylord, B., Reed, D., Raimondi, P., Washburn, L. & McLean, S. A physically based model of macroalgal spore dispersal in the wave and current-dominated nearshore. *Ecology* **83**, 1239–1251 (2002).
- Araki, K., Shimatani, K. & Ohara, H. Dynamics of distribution and performance of ramets constructing genets: a demographic–genetic study in a clonal plant, *Convallaria keiskei*. *Ann. Bot. (Lond.)* **104**, 71–79 (2009).
- Narbonne, G. M. & Gehling, J. G. Life after snowball: the oldest complex Ediacaran fossils. *Geology* **31**, 27–30 (2003).
- Peterson, K. J., Waggoner, B. & Hagadorn, J. W. A fungal analog for Newfoundland Ediacaran fossils? *Integr. Comp. Biol.* **43**, 127–136 (2003).
- Fedonkin, M. A. in *Venskaya Sistema 1, istoriko-geologicheskoe i paleontologicheskoe obosnovanie paleontologiya* Vol. 1 (eds Sokolov, B. S. & Ivanovskiy, A. B.) 70–106 (Nauka, 1985).
- Penny, A. M. et al. Ediacaran metazoan reefs from the Nama Group, Namibia. *Science* **344**, 1504–1506 (2014).
- Yuan, X. et al. The Lantian biota: a new window onto the origin and early evolution of multicellular organisms. *Chin. Sci. Bull.* **58**, 701–707 (2013).
- Hua, H., Chen, Z., Yuan, X., Zhang, L. & Xiao, S. Skeletogenesis and asexual reproduction in the earliest biomineralizing animal *Cloudina*. *Geology* **33**, 277–280 (2005).
- Chen, L., Xiao, S., Pang, K., Zhou, C. & Yuan, X. Cell differentiation and germ–soma separation in Ediacaran animal embryo-like fossils. *Nature* **516**, 238–241 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements The Parks and Natural Areas Division, Department of Environment and Conservation, Government of Newfoundland and Labrador, provided permits to conduct research within the Mistaken Point Ecological Reserve in 2010, while the Department of Tourism, Culture and Recreation provided permits for palaeontological research on the Bonavista Peninsula in 2012. Access to both of the aforementioned fossil localities is by scientific research permit only. Contact the relevant Department listed above for further information. This work has been supported by the Natural Environment Research Council (grant numbers NE/I005927/1 to C.G.K., NE/J5000045/1 to J.J.M., NE/L011409/1 to A.G.L. and NE/G523539/1 to E.G.M.), and a Henslow Junior Research Fellowship from the Cambridge Philosophical Society to A.G.L. We thank M. Laflamme for discussions on this manuscript.

Author Contributions E.G.M. conceived the project, collected data on the 'D' and 'E' surfaces and ran the analyses. C.G.K., A.G.L. and J.J.M. collected data on the H14 surface. All authors discussed the results and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G.M. (ek338@cam.ac.uk).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Data collection. Fossil taxa³¹ and spatial positions on the Mistaken Point 'D' and 'E' surfaces (Extended Data Fig. 1a, b, d and e) were recorded using differentiated GPS over 15 days. A fixed GPS transmitter was installed on a headland overlooking the bedding planes, and a portable GPS receiver was used to map 4,496 individual fossils over a total area of 123.7 m² for both surfaces (Supplementary Information section 2.1); the mean accuracy of data points was 0.4 ± 0.06 cm horizontally and 0.82 ± 0.11 cm vertically. The measured position of each specimen represents the mean of five separate GPS readings collected over 5 s; both the accuracy (standard deviation of the five readings) and the associated weather conditions were recorded for each reading. Mann–Whitney tests were used to compare the densities of specimens recorded in differing weather conditions with the null hypothesis that the density should not depend on the weather conditions on the day of data collection.

Fossil positions on the H14 surface, Bonavista Peninsula (Extended Data Fig. 1f), were recorded by tracing the outline of each specimen onto 2 m × 5 m acetate sheets. Cleavage and other geological features were also traced. These data were collected by three people, two holding the sheets in position and the third recording the data. Wind-induced slippage (affecting large-scale spatial relationships (>0.75 m) for three out of five of the sheets) was determined by measuring the differences between cleavage features crossing sheets, yielding a mean accuracy of 1.47 ± 0.26 cm along strike, and 1.53 ± 0.08 cm parallel to dip. Over the 0.5 m distance that the PCFs were calculated, these errors translate to 0.37 ± 0.26 cm along strike and 0.15 ± 0.08 cm parallel to dip—substantially less than the 1 cm cells within which specimen densities were measured to calculate the PCFs (Methods: Testing for non-random spatial distributions). The sheet approach was used to map the H14 surface because it provided size data more efficiently than direct measurement plus GPS.

Specimen identification. Specimens were recorded as belonging to 1 of 12 taxonomic groups of macrofossils, including two 'bin' groups³², as follows. 1, *Bradgatia*. 2, *Pectinifrons*. 3, *Thectardis*. 4, *F. andersoni* + *F. misrai*. 5, *Charniodiscus spinosus* + *Charniodiscus procerus*. 6, 'Feather dusters'. 7, *Hiemalora*. 8, *Ivesheadiomorphs*³³. 9, Lobate discs. 10, *Charnia* 'A' + *Charnia* 'B'. (*Charnia* 'A' consists of *Beothukis mistakensis*^{34,35} (which dominates the 'E' surface) and *Charnia masoni*. *Charnia* 'B' is now reassigned as *Trepassia wardae*³⁴. Charniid populations on Mistaken Point are dominated by *Beothukis* (only four individuals on the 'E' surface are true *Charnia* species), therefore direct comparison of data from this grouping with those from other taxonomic groups should be undertaken with caution.) 11, 'Holdfast discs' (all discoidal specimens of uncertain affinity, with or without associated stems, lacking sufficient detail to identify the taxon). 12, 'Other species' (rare forms that do not fall into any of the other groups; for example, *Hapsidophyllas*).

Retrodeformation. The tectonically distorted data from the Mistaken Point surfaces were retrodeformed by returning elongated holdfast discs to a circular outline⁶. The 'D' surface (based on 13 specimens) showed a deformation factor of 1.35 ± 0.11 ($R^2 = 0.92$), and the 'E' surface (based on 12 specimens) 1.71 ± 0.08 ($R^2 = 0.754$), both within the previously measured range¹⁰. In the absence of any obviously directional distortion or suitable deformation indicators, measurements taken from the H14 locality were not adjusted in this fashion.

Data collection bias. The impact of mechanical weathering on the 'D' and 'E' surfaces was investigated by modelling the fossil distributions as heterogeneous Poisson processes. Fossils were originally covered in a thin layer of volcanic tuff, which has since been partly weathered away to expose the bedding planes, potentially inducing bias. If the density of a particular taxon is correlated to modern weathering features, then such processes are likely to be masking the true palaeontological distribution of the fossils. Initial data exploration and residual analysis of weathering effects was performed in R³⁶ using the package *spatstat*³⁷. Four covariates, corresponding to four potential erosion sources, were investigated as follows. (1) Across the bedding plane (south to north) x , which is differentially eroded by cliff fall and water runoff from a small stream on the northern side for the 'E' surface. (2) Along the three bedding planes (west to east) y , which are subject to differential erosion from wave action. (3) The southwestern corner xy , which is the first point of contact for most waves on the 'E' surface. (4) The height of the fossils above the troughs of the tectonic ripples on the 'D' and 'E' surfaces h , which is an inverse proxy for ash coverage.

On each bedding plane, and for each parameter, the spatial density of fossils in relation to the parameter was plotted, along with the best-fit quadratic line. This best-fit line was then used to model the change of density compared with the covariant. The inhomogeneous models were tested primarily on the

non-retrodeformed data (since retrodeformation may mask any aggregation due to preservational bias), and verified by conducting similar tests on the retrodeformed data. Two different methods were used to compare the different inhomogeneous models: Kolmogorov–Smirnov tests were performed on quadrats of the data to investigate the distributions relative to the four covariates³⁸ (Supplementary Table 1), then the model fit was assessed using the model residuals^{5,8}. Model residuals assessed the fit of the model to the data by plotting Q–Q and smoothed residual plots. If the observed line in the Q–Q plot fell outside two standard deviations of the model, the model was rejected^{5,8}. Akaike information criterion values³⁹ were used to compare the relative quality of the statistical models that fitted the data.

Bias generated by differing light conditions was tested by comparing fossil densities in areas either side of a specific grid line: on one side the fossils were mapped under optimal conditions (clear skies and low-raking light), and on the other under sub-optimal conditions (cloud-cover with limited shadow-cast). The extent of ash erosion was similar across this grid line, so any measured differences in fossil density are likely to be an artefact of observer bias resulting from different weather conditions. The densities were then compared using a Mann–Whitney test.

Testing for non-random spatial distributions. Initial data exploration, inhomogeneous Poisson modelling and residual analysis were performed in R³⁷ using the package *spatstat*³⁸. Programita⁴⁰ was used to find distance measures and to perform aggregation model fitting (described in detail in refs 40–42). PCFs were used to assess which spatial distributions did not exhibit CSR, where the PCF value reflects how many times more likely the distribution seen is aggregated (or segregated) compared with CSR, as follows. (1) A distribution map was plotted for individual taxa, with the surfaces split into a grid of 1 cm × 1 cm cells, within which the population density was calculated. (2) The smoothed PCF was calculated with smoothing dependent on number of specimens for each taxon. A 3-cell smoothing was applied for *Fractofusus* (D, E and H14), 5 cells for *Charniodiscus* (E), *Pectinifrons* (D), *Bradgatia* (D and E) and *Primocandelabrum* (E), and 15 cells for *Thectardis* (E) and charniids (D and E). (3) Ninety-nine simulations were run for each taxon on a homogeneous background to generate simulation envelopes around the random (PCF = 1). Ninety-nine simulations were run (instead of 100, for example) so that the p_d values could be measured in 0.01 increments. (4) Values for p_d were calculated using Diggle's goodness-of-fit test¹⁵.

For those taxa found to exhibit excursions outside the simulation envelope, four types of process were then fitted to the data: heterogeneous Poisson process, Thomas single cluster processes on both homogeneous and heterogeneous backgrounds, and Thomas double cluster process. The resulting models were then compared to find the best model for each taxon.

Complexities of assessing model fit. Testing for significance with spatial point data is more complicated than for classical statistics owing to lack of independence and variety of point pattern distributions⁴. Monte Carlo simulations provide a good assessment, but the simulation envelope does not necessarily correspond to a confidence interval¹⁵, and runs the risk of type 1 error if the observed PCF falls near the edge of the simulation envelope¹⁵. The size of simulation envelopes depends on the sample size, so that smaller sample sizes (such as the H14 large size-class of 350 specimens) have a relatively large simulation envelope in contrast to the 'D', 'E' and H14 surfaces (all >1,000 specimens). A comparatively large simulation envelope reduces the likelihood that the null model (such as CSR) is rejected. Consequently, hypothesis testing needs to be further supplemented. We used Diggle's goodness-of-fit test, which is a single test statistic¹⁵ (p_d) representing the total squared deviation between the observed pattern and the theoretical result across the studied distances. This test statistic was used in conjunction with visual inspection of Monte Carlo simulations for two reasons. First, p_d does not strictly test whether a model should be accepted or rejected, but whether the PCFs for the observed data are within the range of the stochastic realization of the model⁴³. Second, p_d depends on the range over which it is calculated. For example, the model that best describes the 'E' surface data has $p_d = 0.56$, which may appear low. However, inspection of the PCF (Extended Data Fig. 4b) shows a very close fit to the double Thomas cluster model above 2 cm (Extended Data Table 2). The finite size of *Fractofusus* is reflected in the lower PCF values at small distances, so the model is only fit above 2 cm.

Interpreting ecological processes from spatial point patterns is imprecise. Different processes can produce similar spatial patterns^{8,18,44,45}, with the complex interplay of intra- and interspecific interactions affecting organismal distributions^{46–48}. Even so, application of complementary statistical techniques, such as PCFs combined with comparisons of inhomogeneous Poisson and Thomas cluster models, and nearest neighbour distance analysis, offers the most effective means of teasing out the underlying ecological processes.

Model fitting. If a taxon was not randomly distributed on a homogeneous background (Extended Data Table 1), the random model on a heterogeneous

background was tested. Six different heterogeneous backgrounds were generated, as follows (Extended Data Table 4). (1) The first heterogeneous background was created from the density map of the taxon under consideration, being defined by a circle of radius r over which the density is averaged throughout the sample area. Density maps were formed using estimators within the range of $0.1\text{ m} < r < 1\text{ m}$, and the radius corresponding to the best-fit model was used. (2) The second heterogeneous background was created from density maps of all specimens on each surface combined. (3) The third to sixth heterogeneous backgrounds were created from the four separate density maps of *Fractofusus*, *ivesheadiomorphs*, *Charniodiscus* and *Primocandelabrum*.

This procedure follows that used to test for a non-random distribution on a homogeneous background (Methods: Testing for non-random spatial distributions), except at point 3, where the homogeneous background on which the taxa were simulated is replaced by a heterogeneous one. If excursions outside the simulation envelopes for both homogeneous and heterogeneous Poisson models remained, then cluster models were fitted to the data. For each non-random taxon, univariate cluster models were fitted as follows (Extended Data Table 1). (1) The PCF and L-function⁴⁹ of the observed data were found. Both measures were calculated to ensure that the best-fit model was not optimized towards only one distance measure, and thus encapsulated all spatial characteristics. (2) Best-fit Thomas cluster processes⁵⁰ were fitted to the two functions where $\text{PCF} > 1$. The best-fit lines were not fitted to fluctuations around the random line of $\text{PCF} = 1$, so as to aid good fit about the actual aggregations and to limit fitting of the model about random fluctuations. Programita used the minimal contrast method^{18,15} to find the best-fit model⁸. (3) If the model did not describe the observed data well, the lines were refitted using just the PCF. If that fit was also poor, then only the L-function was used. (4) Ninety-nine simulations of this model were generated to create simulation envelopes, and the fit checked using the O-ring statistic⁴⁰. (5) The value of p_d was calculated over the model range. Very small-scale segregations (under 2 cm) were not included in the model fitting, since they probably represented the finite size of the specimens, and the lack of specimen overlap. (6) If there were no excursions outside the simulation envelope and the value of p_d was high, then a univariate homogeneous cluster model was interpreted as the best model.

Taxa exhibiting two scales of clustering were modelled as Thomas double cluster processes on a homogeneous background, and as single Thomas cluster processes on a heterogeneous background as follows (Extended Data Table 2). (1) The PCF was plotted and ranges for the two different scales of clustering were found. For example, the small-scale cluster may be $0 < r < 0.5\text{ m}$, and the large-scale cluster $0.5\text{ m} < r < 1.5\text{ m}$. (2) The large-scale cluster model was fitted. (3) The parameters of the large-scale single cluster model were used as parameters for the large-scale clusters of the double cluster model. (4) Nearest-neighbour functions were calculated and compared with the parameters of the different-scaled clusters to test for nestedness.

Comparison between and within taxa. To assess whether *Fractofusus* spatial distributions could be similarly modelled on all three bedding planes, the best-fit model from each surface was fitted to the other two surfaces. Simulation envelopes and values of p_d were used to evaluate the fit (Fig. 2b and Extended Data Table 3).

The uniqueness of each taxon's spatial distribution was assessed by fitting the best-fit models for high-abundance taxa (*Fractofusus*, *Charniodiscus*, *Primocandelabrum*) onto each other. Low-abundance taxa (*charniids*, *Thectardis*) were excluded from this comparison because they yielded large simulation envelopes, and consequently very different models fitted within the generated envelopes. **Size-classes.** The *Fractofusus* size data from the H14 surface allowed investigation of interactions between *Fractofusus* specimens of different sizes. To determine whether there was a dependency between spatial distribution and specimen size, the mark correlation function was calculated^{17,51} and compared with the simulation envelope produced from 99 Monte Carlo simulations (Extended Data Fig. 3a).

The most objective way to resolve the number and range of size-classes in a population is by fitting size–frequency distribution data (the natural log of the variables length, width and the bivariate case of length multiplied by width) to various models, followed by comparison of (logarithmically scaled) Bayesian information criterion values⁷, which we performed in R using the package MCLUST⁵². The number of populations thus identified was then used to define the most appropriate size-classes. A Bayesian information criterion value difference of > 10 corresponds to a 'decisive' rejection of the hypothesis that two models are the same, whereas values < 6 indicate only a weak rejection of the similarity of the models.⁵³

Once defined, the spatial distributions for each size-class were analysed using the techniques described in the 'Model fitting' section. Although it was necessary to set firm boundaries for each size-class, the populations were normally

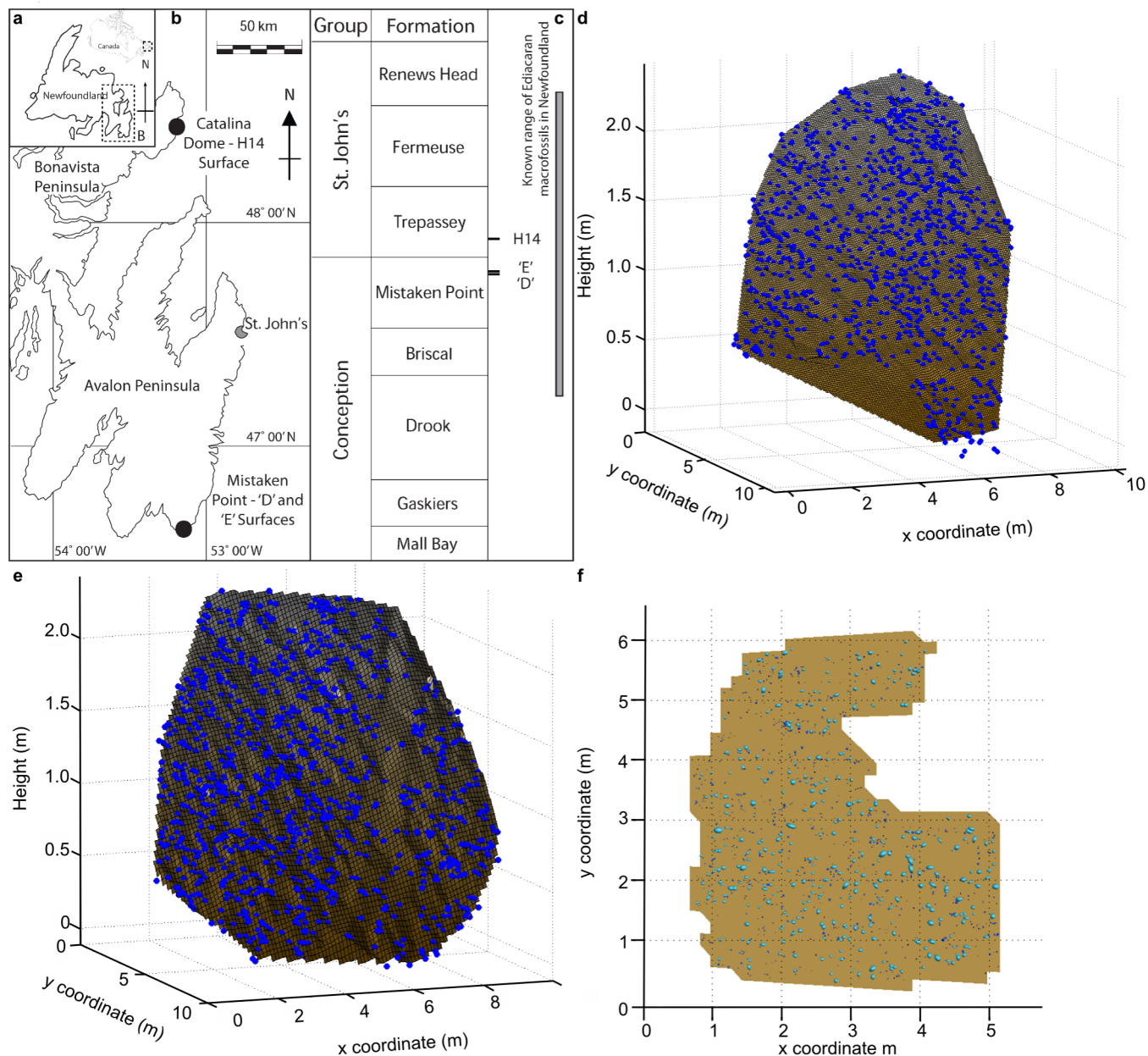
distributed and therefore overlapped. As a result, the largest individuals of the small population were grouped within the middle size-class, while some of the smallest of the medium population were included within the small size-class.

Using this information, further analyses were performed to consider the spatial relationships between individual size-classes. The PCF between each size-class was determined, the best-fit shared parent and LCMs were fitted (Extended Data Table 5 and Extended Data Fig. 4c, d) and the fit of each model assessed. Shared parent models describe the pattern when two clustered size-classes both cluster around an (un-defined) point or area, such as the effect of two different heterogeneous soil effects on tree growth⁵³. LCMs arise when one size-class clusters around another size-class, for example when tree saplings cluster around their parent⁵⁴. The procedure behind model fitting was similar to the single group/size-class (univariate) case; however, instead of single groupings, two size-class (bivariate) PCFs were used instead, as follows. (1) The best-fit Thomas cluster processes were found for each of the size-classes (shared parent and LCM). (2) The single size-class parameters for each size were input to the model classes (both models for the shared parent model and only the parent model for LCM). (3) The best-fit model was fitted to the PCF and L-functions. (4) If the model fit was poor (errors > 0.025), the model was fitted using the PCF and then the L-function. If neither were a good fit, then the spatial scale that the model was fitted to was reduced, so that a good model could be found for at least part of the spatial scale. (5) The model was checked using 99 Monte Carlo simulations, p_d values, and by comparing the univariate parameters with the bivariate model parameters.

Isotropy analysis. To assess whether non-random behaviour was stronger in any particular direction (that is, it exhibited isotropy), density plots of the K-measure^{5,55} were used to calculate the normalized density in each direction around each point (Fig. 3 and Extended Data Fig. 3c), where normalized values of 1 indicate random distribution (homogeneous Poisson process) compared with < 1 (segregation) and > 1 (aggregation). The plots were produced by calculating the average of all the vectors that joined all pairs of points over different realizations of the point process. Each point in turn was positioned on the plot centre, then a vector drawn to every point. The resulting vector scatter plot (also known as a Fry plot) was then smoothed using a Gaussian smoothing kernel¹⁵.

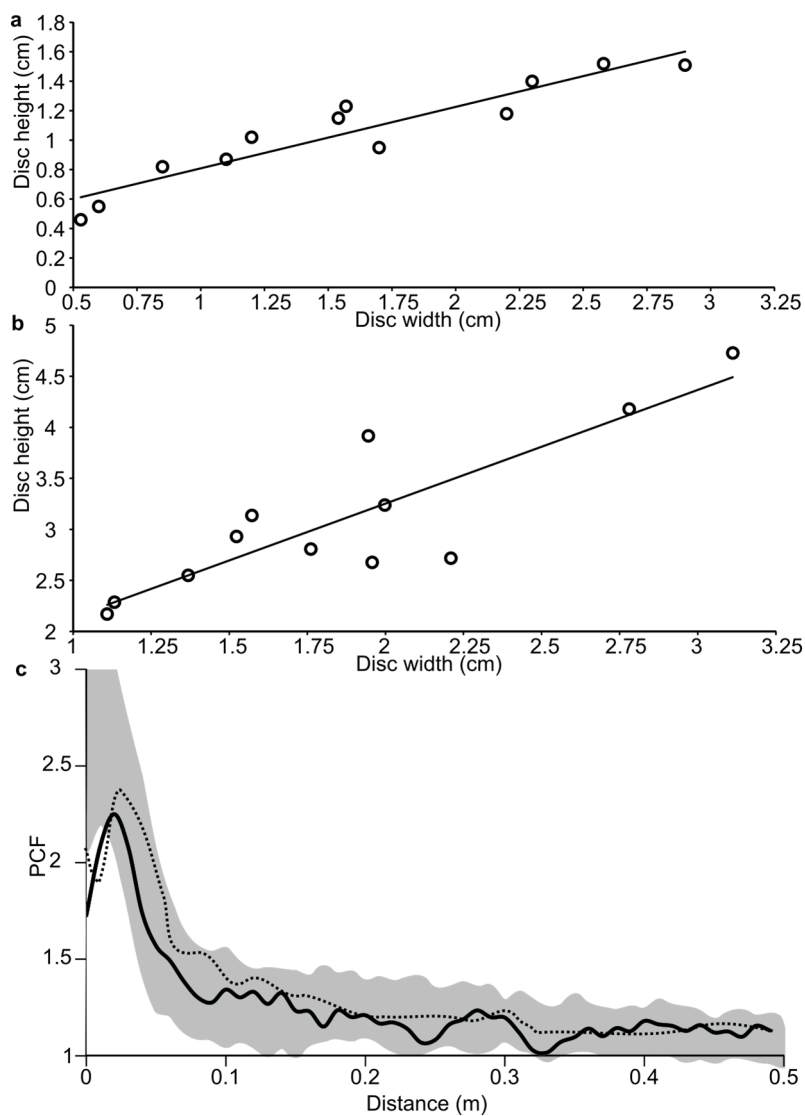
31. Clapham, M. E. in *Quantifying the Evolution of Early Life* (eds Laflamme, M., Schiffbauer, J. D. & Dornbos, S. Q.) 3–21 (Springer, 2011).
31. Shen, B., Dong, L., Xiao, S. & Kowalewski, M. The Avalon explosion: evolution of Ediacara morphospace. *Science* **319**, 81–84 (2008).
33. Liu, A. G., McIlroy, D., Antcliffe, J. B. & Brasier, M. D. Effaced preservation in the Ediacara biota and its implications for the early macrofossil record. *Paleontology* **54**, 607–630 (2011).
34. Narbonne, G. M., Laflamme, M., Greentree, C. & Trusler, P. Reconstructing a lost world: Ediacaran rangeomorphs from Spaniard's Bay, Newfoundland. *J. Paleontol.* **83**, 503–523 (2009).
35. Brasier, M. D. & Antcliffe, J. B. Evolutionary relationships within the Avalonian Ediacara biota: new insights from laser analysis. *J. Geol. Soc. Lond.* **166**, 363–384 (2009).
36. R Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2013).
37. Baddeley, A. & Turner, R. Spatstat: an R package for analyzing spatial point patterns. *J. Stat. Softw.* **12**, 1–42 (2005).
38. Berman, M. Testing for spatial association between a point process and another stochastic process. *Appl. Stat.* **35**, 54–62 (1986).
39. Baddeley, A., Rubak, E. & Møller, J. Score, pseudo-score and residual diagnostics for spatial point process models. *Stat. Sci.* **26**, 613–646 (2011).
40. Wiegand, T. & Moloney, K. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **104**, 209–229 (2004).
41. Wiegand, T., Kissling, W., Cipriotti, P. & Aguiar, M. Extending point pattern analysis for objects of finite size and irregular shape. *J. Ecol.* **94**, 825–837 (2006).
42. Wiegand, T., Moloney, K., Naves, J. & Knauer, F. Finding the missing link between landscape structure and population dynamics: a spatially explicit perspective. *Am. Nat.* **154**, 605–627 (1999).
43. Loosmore, N. B. & Ford, E. D. Statistical inference using the G or K point pattern spatial statistics. *Ecology* **87**, 1925–1931 (2006).
44. Wiegand, T. & Moloney, K. A. *Handbook of Spatial Point-Pattern Analysis in Ecology* (CRC, 2013).
45. Levin, S. A. in *Ecological Time Series Vol. 2* (eds Powell, T. M. & Steele, J. H.) 277–326 (Springer, 1995).
46. McIntire, E. J. & Fajardo, A. Beyond description: the active and effective way to infer processes from spatial patterns. *Ecology* **90**, 46–56 (2009).
47. Fragoso, J. M., Silvius, K. M. & Correa, J. A. Long-distance seed dispersal by tapirs increases seed survival and aggregates tropical trees. *Ecology* **84**, 1998–2006 (2003).
48. Russo, S. E. & Augspurger, C. K. Aggregated seed dispersal by spider monkeys limits recruitment to clumped patterns in *Virola calophylla*. *Ecol. Lett.* **7**, 1058–1067 (2004).
49. Besag, J. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. B* **36**, 192–236 (1974).

50. Thomas, M. A generalization of Poisson's binomial limit for use in ecology. *Biometrika* **36**, 18–25 (1949).
51. Grabarnik, P., Myllymäki, M. & Stoyan, D. Correct testing of mark independence for marked point patterns. *Ecol. Model.* **222**, 3888–3894 (2011).
52. Fraley, C. & Raftery, A. E. MCLUST version 3: an R package for normal mixture modeling and model-based clustering (Department of Statistics, Washington University, 2006).
53. Fraley, C. & Raftery, A. E. Bayesian regularization for normal mixture estimation and model-based clustering. *J. Classific.* **24**, 155–188 (2007).
54. Pélissier, R. & Goreaud, F. A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. *J. Veg. Sci.* **12**, 99–108 (2001).
55. Stoyan, D., Kendall, W. S. & Mecke, J. *Stochastic Geometry and its Applications* 2nd edn (Springer, 1995).



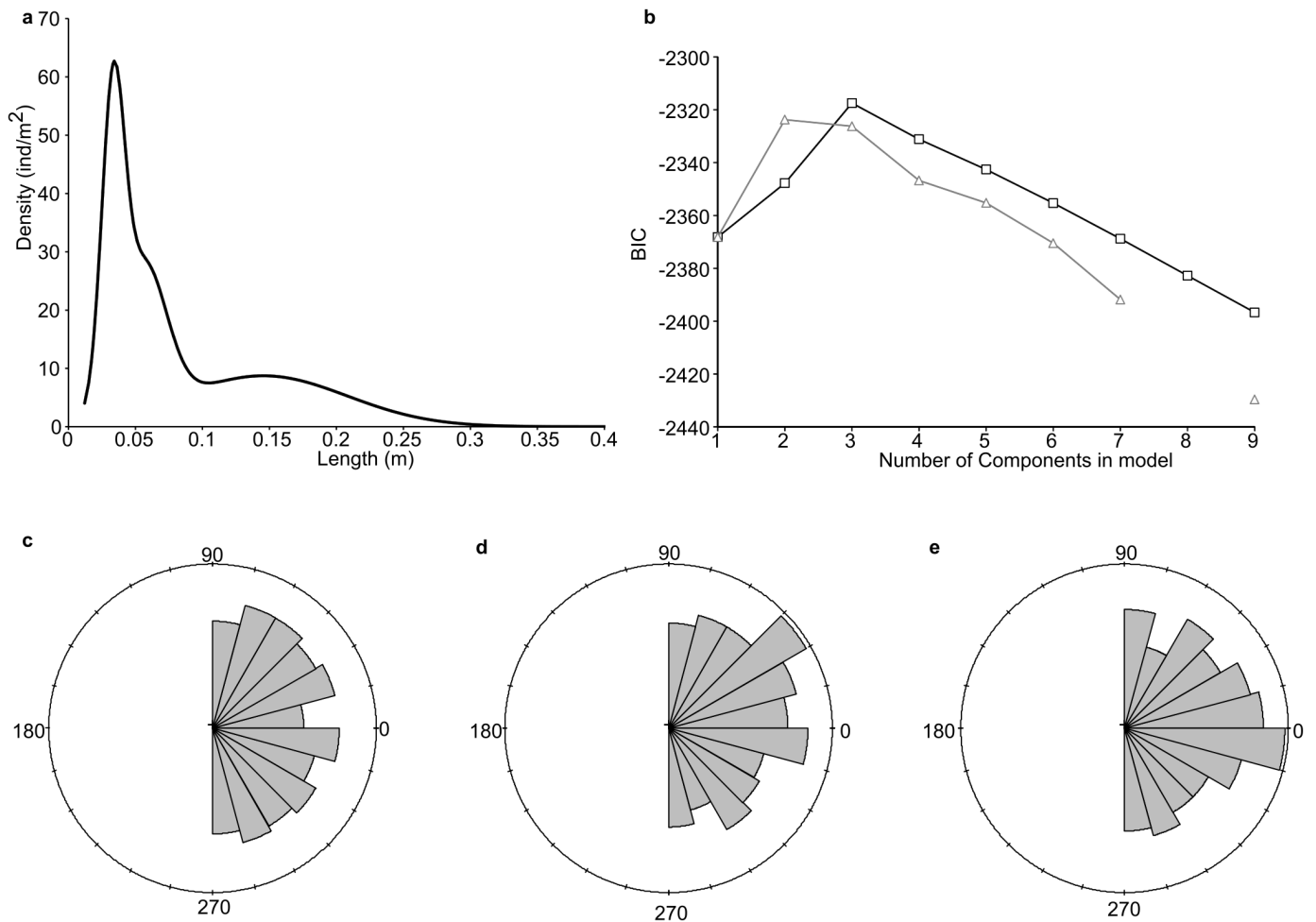
Extended Data Figure 1 | Map and simplified stratigraphic column showing the position of studied bedding planes with bedding plane maps of *Fractofusus*. **a**, Newfoundland, eastern Canada. Dashed area indicates region of interest in **b**. **b**, The Avalon and Bonavista Peninsulas, eastern Newfoundland. Locations of the bedding planes are indicated. **c**, Stratigraphic column (not to scale) compiled from the Avalon and Bonavista Peninsulas; lithological units in each region are treated as correlative in

this study, but work is continuing to determine the validity of this assumption. The 'E' surface at Mistaken Point has been dated to 565 ± 3 Ma (ref. 12). There are currently no available radiometric dates from the Bonavista Peninsula. **d–f**, Maps of *Fractofusus* positions on the 'D' surface (**d**), the 'E' surface (**e**) and the H14 surface (**f**). In **e** the largest specimens are in light blue, medium specimens in mid-blue and smallest specimens in dark blue.



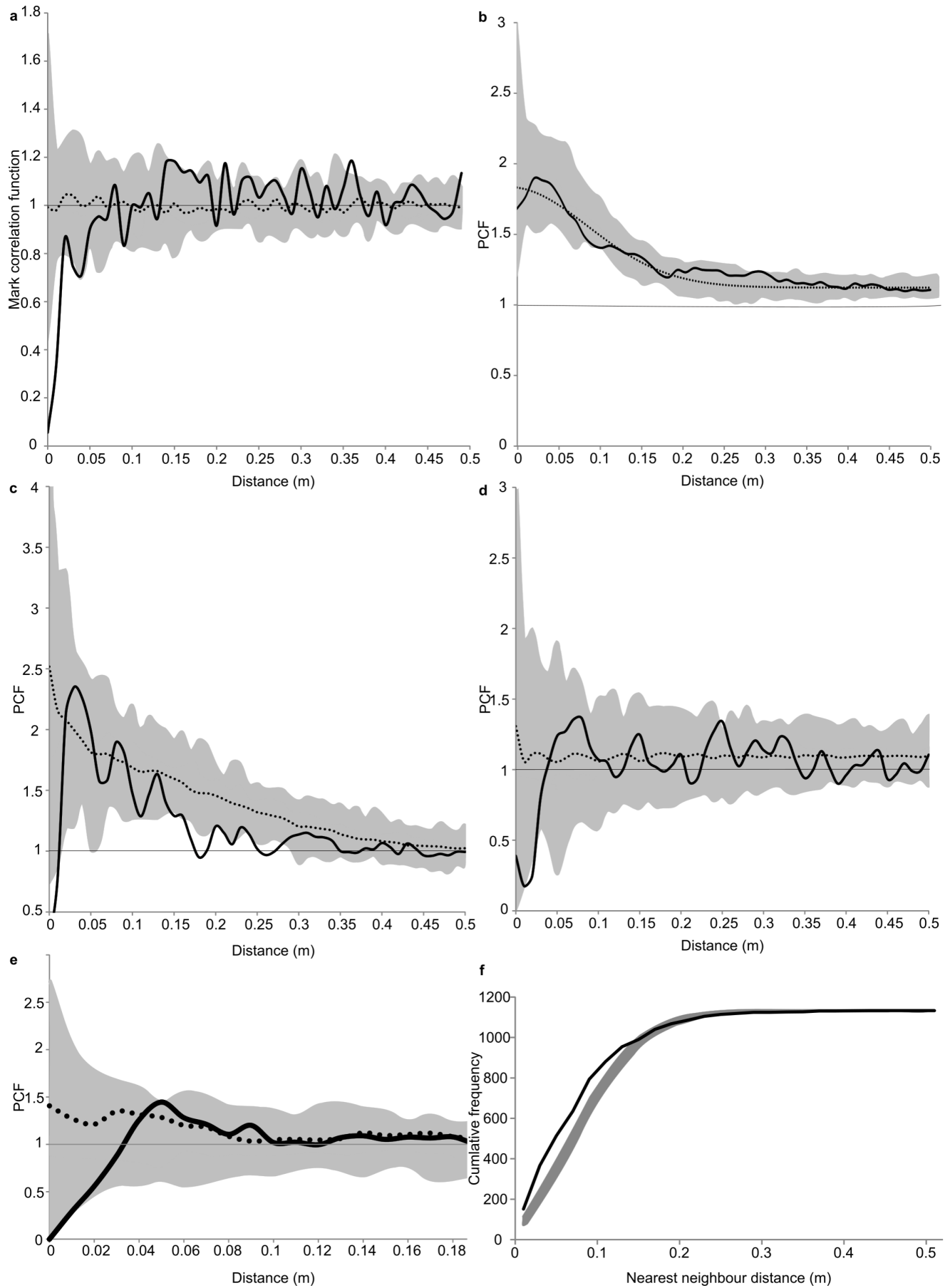
Extended Data Figure 2 | Retrodeformation calculations on the Mistaken Point surfaces. **a, b,** Plots of the lengths versus widths of discs from the 'D' surface, Mistaken Point (**a**), and the 'E' surface, Mistaken Point (**b**). The gradient of the line defines the retrodeformation factor, which for the 'D' surface is 1.35 ± 0.11 ($R^2 = 0.92$) and for the 'E' surface is 1.71 ± 0.08

($R^2 = 0.75$). **c,** *Fractofusus* PCF on the 'E' surface with (solid line) and without (dashed line) retrodeformation. The grey shaded area depicts the boundary of 99 Monte Carlo simulations for the model which provided the best-fit model to the retrodeformed data, which has a good fit on the non-retrodeformed data ($p_d = 0.60$).



Extended Data Figure 3 | Size distribution analysis of *Fractofusus* for the H14 surface. **a**, Size–frequency distributions for *Fractofusus*, ($n = 1,214$); **b**, the results of the Bayesian information criterion^{52,53} (univariate data). Squares and triangles correspond to models assuming equal and unequal variance, respectively. High values of the Bayesian information criterion correspond to a good model fit, so the best-fit model is a three-component

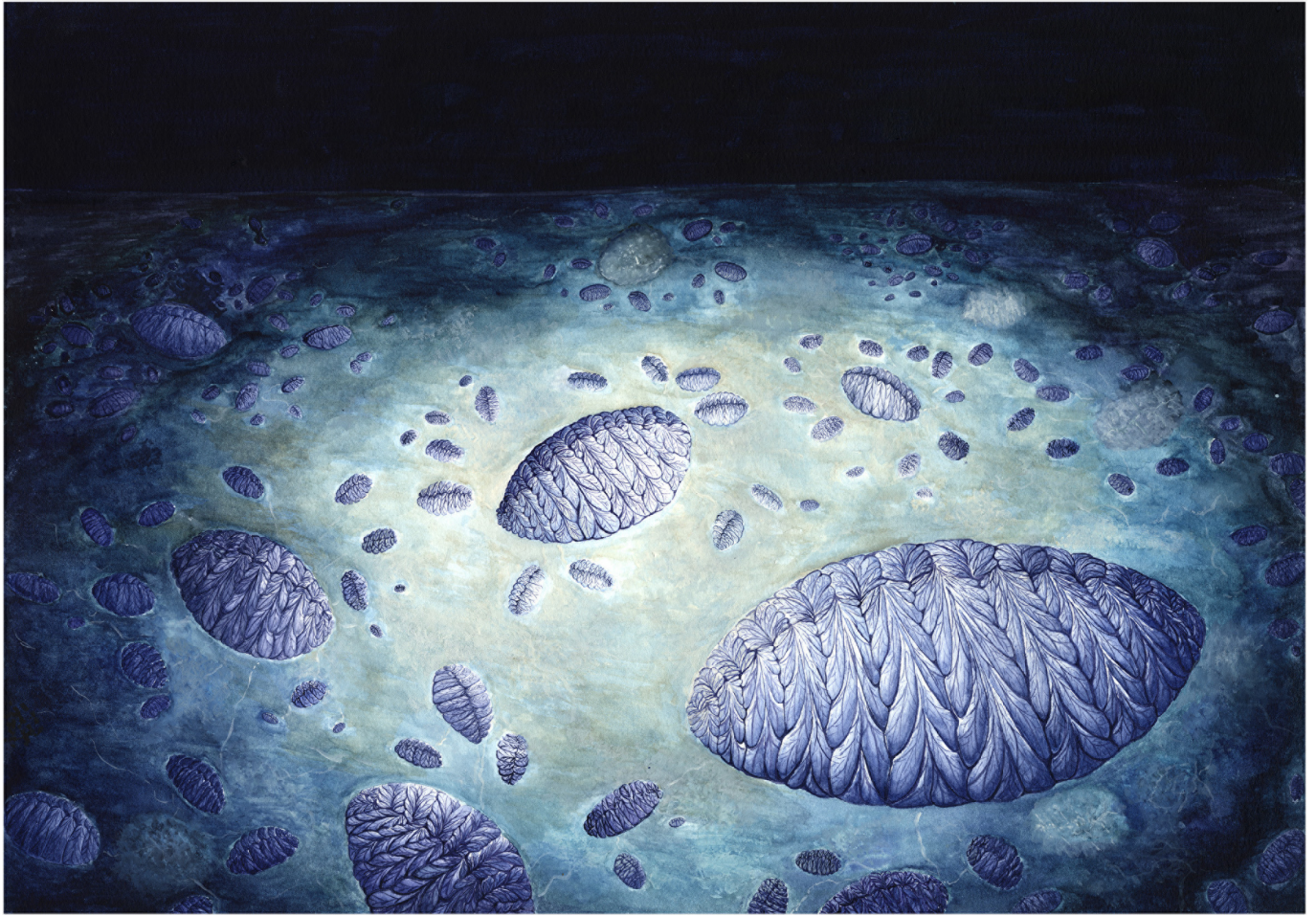
equal variance model using log-normalized length data. **c–e**, Rose diagrams plotting the directional orientation of the different size-classes of *Fractofusus* on the H14 surface showing large size-class (<11.0 cm, $n=350$) (**c**), intermediate size-class (5.5–11.0 cm, $n=310$) (**d**) and small size-class (<5.5 cm, $n=554$) (**e**). The angles of the *Fractofusus* central axis are relative to north (0°). There is no strong orientation preference for any of the size-classes.



Extended Data Figure 4 | Distance measures for the size data from H14

surface. For all plots, the x axis is the inter-point distance between organisms (in metres). **a**, Mark correlation function⁵, where 1 corresponds to a lack of correlation of size, such that *Fractofusus* size is independent and identically distributed. A value of <1 corresponds to a positive dependency (in contrast to PCF) and >1 corresponds to a negative dependency. Small *Fractofusus* on the H14 surface (<0.3 cm) are more likely to be found near each other than expected by random. **b**, The 'E' surface PCF (solid line) showing the model that fits the data best, a double Thomas cluster model (dotted line, $p_d = 0.56$), and

the simulation envelope for 99 Monte Carlo simulations (grey shaded area). **c, d**, PCF for the best-fit models for the bivariate size-classes of *Fractofusus* on the H14 surface showing LCMs for small with medium size-classes ($p_d = 0.74$) (**c**) and LCMs for medium with large size-classes ($p_d = 0.66$) (**d**). **e**, The PCF of the largest size-class of H14 (solid line), showing the CSR Monte Carlo simulation envelope in grey, with the 'D' surface PCF (dotted line, $p_d = 0.56$). **f**, Nearest neighbour distances (solid line, $p_d = 0.01$) with CSR Monte Carlo simulation envelope in grey.



Extended Data Figure 5 | Artistic reconstruction of *Fractofusus* on the H14 surface, Bonavista Peninsula. The bottom right features a large *Fractofusus* around which there are five to eight medium specimens clustered. Each of

the medium specimens also has small specimens clustered around them. The small specimens therefore form an independent double cluster pattern, namely clusters of clusters. Artwork by C.G.K.

Extended Data Table 1 | Best-fit univariate cluster models

	Model Fit				p_d Values				
	σ	100 ρ	Mean number in cluster	Number of Clusters	Cluster on homogeneous Background	Cluster on heterogeneous Background	Heterogeneous Poisson Model	Homogeneous Poisson Model	Error function for best fit model
Taxon									
<i>Fractofusus</i> (D)	4.312	0.638	3	338	0.77	0.21	0.51	0.01	0.00346
<i>Charniodiscus</i> (E)	2.133	1.667	10	16	0.41	0.14	0.13	0.01	0.02438
Charniid (E)	6.841	0.11	14	7	0.60	0.11	0.90	0.01	0.00697
<i>Primocandelabrum</i> (E)	5.616	0.329	11	19	0.28	0.18	0.09	0.01	0.00057
<i>Fractofusus</i> (E)	11.842	0.39	25	43	0.32	0.35	0.03	0.01	0.00064
<i>Thectardis</i> (E)	1.835	2.003	4	20	0.91	0.07	0.44	0.02	0.00290
<i>Fractofusus</i> (H14)	7.906	0.192	2	728	0.25	0.01	0.03	0.01	0.00276
Small size (H14)	3.291	0.203	7	771	0.17	0.02	0.01	0.01	0.00021
Medium size (H14)	4.377	0.188	4	714	0.86	0.15	0.01	0.01	0.01151
Large size (H14)	NA	NA	NA	NA	NA	NA	NA	0.30	NA
<i>Ivesheadiomorphs</i> (E)	NA	NA	NA	NA	NA	NA	NA	0.21	NA
Lobate Discs (E)	NA	NA	NA	NA	NA	NA	NA	0.11	NA
<i>Bradgatia</i> (E)	NA	NA	NA	NA	NA	NA	NA	0.44	NA
<i>Pectiniifrons</i> (D)	NA	NA	NA	NA	NA	NA	NA	0.28	NA
<i>Bradgatia</i> (D)	NA	NA	NA	NA	NA	NA	NA	0.51	NA
<i>Ivesheadiomorphs</i> (D)	NA	NA	NA	NA	NA	NA	NA	0.13	NA
Charniid (D)	NA	NA	NA	NA	NA	NA	NA	0.10	NA

For the heterogeneous backgrounds, the moving window radius is 0.5 m using the same taxon density as the taxon being modelled. A value of $p_d = 1$ corresponds to a perfect fit of the model on the data, while $p_d = 0$ corresponds to no fit. The error function of the best-fit model gives the fraction of the total sum of squares for the transformed empirical PCF that is not explained by the model.

Extended Data Table 2 | Best-fit univariate double cluster models

Taxon	Large Scale Clusters					Small Scale Clusters				
	NN (m)	σ_1	$100p_1$	Mean number of small clusters	Number of clusters	σ_2	$100p_2$	Mean Number in cluster	p_d value	Error function on Model Fit
<i>Charniodiscus</i> (E)	0.23	11.015	0.395	22	23	4.184	0.203	8	0.63	0.00031
<i>Fractofusus</i> (E)	0.17	12.191	0.392	12	23	3.675	0.470	3	0.56	0.00147
<i>Primocandelabrum</i> (E)	0.27	7.123	0.282	14	16	2.109	4.698	4	0.72	0.00036
<i>Fractofusus</i> (H14)	0.12	11.859	0.166	6	24	3.953	1.002	8	0.89	0.00063
Small size class (H14)	0.13	9.873	0.078	6	19	1.646	0.808	5	0.72	0.00021

Large-scale clusters are determined for the univariate cluster then input into the model, and the small-scale clusters are determined in the double cluster analysis. A value of $p_d = 1$ corresponds to a perfect fit of the model on the data, while $p_d = 0$ corresponds to no fit. NN denotes the mean nearest neighbour distance.

Extended Data Table 3 | Best-fit double Thomas cluster models fitted onto other taxa

		Best-fit Model p_d value				
	Taxon	<i>Fractofusus</i>	<i>Fractofusus</i>	<i>Fractofusus</i>	<i>Charniodiscus</i>	<i>Primocandelabrum</i>
	Surface	H14	D	E	E	E
Taxon	Surface					
<i>Fractofusus</i>	H14	0.89	0.01	0.06	NA	NA
<i>Fractofusus</i>	D	0.01	0.77	0.01	NA	NA
<i>Fractofusus</i>	E	0.51	0.01	0.56	0.16	0.07
<i>Charniodiscus</i>	E	NA	NA	0.02	0.68	0.01
<i>Primocandelabrum</i>	E	NA	NA	0.51	0.01	0.72

A value of $p_d = 1$ corresponds to a perfect fit of the model on the data, while $p_d = 0$ corresponds to no fit. Note that while these numbers may seem low (such as the 'E' surface fit), they need to be considered in context of the PCF graph (Extended Data Fig. 4b), which clearly shows a good fit to the data, with the small fluctuations of the observed PCF around the model PCF.

Extended Data Table 4 | Best-fit univariate cluster models on heterogeneous backgrounds for 'E' surface taxa

Taxon	Fitted Model Parameters					Heterogeneous Cluster Models p_d values									
	Spatial Scale	σ	100p	Mean number in Cluster	Number of Clusters	C	CH _{all}	CH _{Ive}	CH _{Frac}	CH _{Cha}	CH _{Primo}	H _{Bra}	H _{Lob}	H _{The}	H _{Char}
<i>Charniodiscus</i>	L	11.015	0.395	22	23	0.11	0.33	0.14	0.16	0.02	0.04	NA	NA	NA	NA
<i>Charniodiscus</i>	S	2.856	2.880	6	167	0.09	0.22	0.03	0.05	0.04	0.14	NA	NA	NA	NA
<i>Fractofusus</i>	L	22.556	0.114	34	34	0.08	0.20	0.03	0.13	0.02	0.01	NA	NA	NA	NA
<i>Fractofusus</i>	S	3.662	2.169	5	726	0.01	0.01	0.01	0.01	0.01	0.01	NA	NA	NA	NA
<i>Primocandelabrum</i>	L	7.123	0.282	14	16	0.16	0.45	0.41	0.42	0.14	0.18	NA	NA	NA	NA
<i>Primocandelabrum</i>	S	2.090	1.057	4	242	0.05	0.20	0.18	0.17	0.16	0.07	NA	NA	NA	NA
Charniid	Random	NA	NA	NA	NA	0.60	0.02	0.04	0.08	0.03	0.04	0.18	0.08	0.12	0.90

Univariate clusters, either fitted to the small scale (S) or large scale (L) were modelled on different backgrounds defined by the density map of all taxonomic groups, or random for charniids. C, Thomas cluster on homogeneous backgrounds; CH, Thomas cluster on heterogeneous backgrounds. For the heterogeneous backgrounds, the moving window radius is 0.5 m since that radius produced the best fit for charniids: heterogeneous cluster model on a background density constructed from all species (CH_{all}), ivesheadiomorphs (CH_{Ive}), *Fractofusus* (CH_{Frac}), *Charniodiscus* (CH_{Cha}) and *Primocandelabrum* (CH_{Primo}). CSR on heterogeneous background (H), *Bradgatia* (H_{Bra}), lobate discs (H_{Lob}), *Thectardis* (H_{The}) and charniids (H_{Char}). A value of $p_d = 1$ corresponds to a perfect fit of the model on the data, while $p_d = 0$ corresponds to no fit. The H14 surface did not possess enough non-*Fractofusus* specimens to perform similar analyses. NA, not applicable.

Extended Data Table 5 | Models for bivariate analysis between different size-classes of *Fractofusus* on the H14 surface

Size class 1	Size class 2	Model type	σ_1	$100\rho_1$	σ_2	$100\rho_2$	σ_t	σ_{best}	$100\rho_{\text{best}}$	Number in Cluster	Number of Clusters	p_d	Error on Fit
Small	Medium	SP	2.829	0.3599	5.714	0.2359	4.51	7.857	0.1372	16	896	0.03	0.00336
Small	Medium	LCM	2.829	0.3599	5.913	0.1027	NA	5.714	0.2359	11	23	0.74	0.00172
Medium	Large	SP	5.913	0.1027	2.836	0.4054	3.7	2.428	1.0128	5	1018	0.01	0.01087
Medium	Large	LCM	5.913	0.1027	2.836	0.4054	NA	1.512	0.1602	5	81	0.66	0.00771
Small	Large	SP	2.829	0.3599	2.836	0.4054	2.83	11.42	0.2337	23	1366	0.03	0.00179
Small	Large	LCM	2.829	0.3599	2.836	0.4054	NA	28.15	0.1844	56	700	0.01	0.00386

A value of $p_d = 1$ corresponds to a perfect fit of the model on the data, while $p_d = 0$ corresponds to no fit. The large size-class was randomly distributed, but was approximated by a cluster model, which was required for input into Programita⁸.

Evolution of endemism on a young tropical mountain

Vincent S. F. T. Merckx^{1,2}, Kasper P. Hendriks^{1,3}, Kevin K. Beentjes¹, Constantijn B. Mennes¹, Leontine E. Becking^{1,4,5}, Katja T. C. A. Peijnenburg^{1,6}, Aqilah Afendy⁷, Nivaarani Arumugam^{7,8}, Hugo de Boer^{1,9,10}, Alim Biun¹¹, Matsain M. Buang¹¹, Ping-Ping Chen¹, Arthur Y. C. Chung¹², Rory Dow¹, Frida A. A. Feijen¹, Hans Feijen¹, Cobi Feijen-van Soest¹, József Geml^{1,2}, René Geurts¹³, Barbara Gravendeel^{1,2,14}, Peter Hovenkamp¹, Paul Imbun¹¹, Isa Ipor¹⁵, Steven B. Janssens¹⁶, Merlijn Jocqué^{17,18}, Heike Kappes^{1,19}, Eyen Khoo¹², Peter Koomen²⁰, Frederic Lens^{1,2}, Richard J. Majapun¹², Luis N. Morgado¹, Suman Neupane²¹, Nico Nieser¹, Joan T. Pereira¹², Homathevi Rahman⁷, Suzana Sabran¹², Anati Sawang⁷, Rachel M. Schwallier¹, Phylau-Soon Shim^{11,‡}, Harry Smit¹, Nicolien Sol¹, Maipul Spait¹¹, Michael Stech¹, Frank Stokvis¹, John B. Sugau¹², Monica Suleiman⁷, Sukaibin Sumail¹¹, Daniel C. Thomas^{1,22,23}, Jan van Tol¹, Fred Y. Y. Tuh¹¹, Bakhtiar E. Yahya⁷, Jamili Nais¹¹, Rimi Repin¹¹, Maklarin Lakim¹¹ & Menno Schilthuizen^{1,2,7}

Tropical mountains are hot spots of biodiversity and endemism^{1–3}, but the evolutionary origins of their unique biotas are poorly understood⁴. In varying degrees, local and regional extinction, long-distance colonization, and local recruitment may all contribute to the exceptional character of these communities⁵. Also, it is debated whether mountain endemics mostly originate from local lowland taxa, or from lineages that reach the mountain by long-range dispersal from cool localities elsewhere⁶. Here we investigate the evolutionary routes to endemism by sampling an entire tropical mountain biota on the 4,095-metre-high Mount Kinabalu in Sabah, East Malaysia. We discover that most of its unique biodiversity is younger than the mountain itself (6 million years), and comprises a mix of immigrant pre-adapted lineages and descendants from local lowland ancestors, although substantial shifts from lower to higher vegetation zones in this latter group were rare. These insights could improve forecasts of the likelihood of extinction and ‘evolutionary rescue’⁷ in montane biodiversity hot spots under climate change scenarios.

In mountainous areas of the humid tropics, steep environmental gradients coincide with high primary productivity and relative climatic stability to sustain large numbers of species, often with striking degrees of endemism at higher elevations^{1,8}. It has therefore been recognized that tropical mountains are biodiversity hot spots of great conservation value¹, especially because endemics on mountain tops are vulnerable to becoming trapped and then annihilated as a result of global warming^{2,3}.

The evolutionary origins of these unique biotas, however, are poorly understood⁴. Like other insular habitats⁹, the endemic biota of an isolated mountain results from complex dynamics among colonization, *in situ* speciation, and local extinction. Each of these factors is dependent on the age and size of the habitat, and on the environmental contrast between the insular habitat and its matrix¹⁰. In the case of a tropical mountain top, an added complication is the fact that climate fluctuations may have widened and restricted the geographic range over which the montane conditions have extended in the past,

meaning that parts of the endemic biota may be relicts, and other components may be novel in character^{4,5}.

Disentangling these possibilities for a single tropical montane biodiversity hot spot requires molecular phylogenetic study of a large number of fauna and flora elements. However, with only few exceptions^{11,12}, evolutionary studies in such hot spots have been limited to single taxa^{13,14}. This precludes broad understanding of the evolutionary and biogeographic origins of an endemic biota as a whole⁴.

We investigated the evolutionary routes to endemism by sampling an entire tropical mountain biota on the UNESCO World Heritage site of Gunung Kinabalu in Sabah, East Malaysia. We demonstrate that most of its unique biodiversity is younger than the mountain itself and comprises a mix of immigrant pre-adapted lineages as well as descendants from local lowland ancestors.

At 4,095 m, Kinabalu is the tallest mountain between the Himalayas and New Guinea. It is a solitary ‘sky island’, having emerged during the Pliocene and early Pleistocene as a granite pluton within the surrounding sandstone of the Crocker Range, the latter having formed much earlier, between the Eocene and the early Miocene¹⁵. Because of the area’s tectonic activity, as well as Pleistocene sea level changes, the exact historical progression of its elevation above sea level is not known, but it is likely that a major rise, even beyond today’s elevation, of Kinabalu, as well as the central spine of the Crocker Range, took place between 6 million years ago and today (for more geological background see Methods). Since the early days of its exploration¹⁶, Kinabalu has been famous for its extremely high biological diversity, especially its richness in endemic species, with endemism proportions reaching 25–30% for some taxa¹⁷.

To unravel the origins of the exceptionally rich Kinabalu biota, we mounted a Malaysian–Dutch expedition in which 47 taxonomists worked at 37 localities, spanning the full range of elevations (Fig. 1). We used Sanger sequencing to sequence one or more fast-evolving loci for 1,852 individuals, belonging to 18 genera representing gastropods, annelids, insects, arachnids, vertebrates, pteridophytes, bryophytes, and angiosperms. We also obtained 3.7 million basidiomycete

¹Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands. ²Institute of Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands. ³Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands. ⁴Wageningen University & Research centre, Marine Animal Ecology Group, PO Box 338, 6700 AH Wageningen, The Netherlands. ⁵Department of Environmental Science, Policy, & Management, University of California Berkeley, 130 Mulford Hall #3114, Berkeley, California 94720, USA. ⁶Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. ⁷Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia. ⁸Faculty of Earth Science, Universiti Malaysia Kelantan, Jeli Campus, Locked bag No.100, 17600 Jeli, Kelantan Darul Naim, Malaysia. ⁹Department of Organismal Biology, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden. ¹⁰Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318 Oslo, Norway. ¹¹Sabah Parks, Lot 45 & 46, Level 1-5, Blok H, KK Times Square, 88806 Kota Kinabalu, Sabah, Malaysia. ¹²Forest Research Centre, Sabah Forestry Department, P.O. Box 1407, 90175 Sandakan, Sabah, Malaysia. ¹³Wageningen University, Department of Plant Sciences, Laboratory of Molecular Biology, 6700AP Wageningen, The Netherlands. ¹⁴University of Applied Sciences Leiden, Zernikedreef 11, 2333 CK Leiden, The Netherlands. ¹⁵Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia. ¹⁶Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium. ¹⁷Royal Belgian Institute of Natural Sciences, Aquatic and Terrestrial Ecology, Vautierstraat 29, 1000 Brussels, Belgium. ¹⁸Rutgers, The State University of New Jersey, Department of Biological Sciences, 195 University Avenue, Boyden Hall, Newark, New Jersey 07102, USA. ¹⁹Zoological Institute, University of Cologne, Zùlpicher Strasse 47b, D-50674 Cologne, Germany. ²⁰Natuurmuseum Fryslân, Schoenmakersperk 2, 8911 EM Leeuwarden, The Netherlands. ²¹EEB Department, University of Connecticut, 75 N. Eagleville Road, Storrs, Connecticut 06269-3043, USA. ²²School of Biological Sciences, University of Hong Kong, Pok Fu Lam Road, Hong Kong, China. ²³Singapore Botanic Gardens, 1 Cluny Road, 259569 Singapore, Republic of Singapore.

‡Deceased.



Figure 1 | Map of the study area. Inset left, location of the study area in the World and in Borneo. Inset middle, detail of the summit trail in Kinabalu Park. The eight expedition stations in Kinabalu Park and Crocker Range Park are

and glomeromycete ITS2 rDNA sequences from soil cores with ion semiconductor sequencing. In addition, we retrieved data from eight previously published single-taxon studies on vascular plants.

We analysed all data within a phylogenetic framework to estimate the times of origin of endemic species, and to determine whether endemic species had descended from local or distant congeners (Methods, Fig. 2, Extended Data Table 1 and Extended Data Figs 1–9). Note that we define ‘endemic’ as restricted to the area in which our expedition took place. Although the present study offers the most comprehensive evolutionary analysis of any mountain biota to date, the taxa covered are, by necessity, an uneven and fragmentary sampling of the full diversity. Nonetheless, we expect that our results are representative for the Kinabalu biota as a whole, as our selected taxa encompass organisms with a wide variety of phylogenetic backgrounds, ecologies, and life history traits.

Similar to Mesoamerican endemic cloud forest seed plants and vertebrates¹¹, our molecular dating results show that the estimated mean stem-node ages of 33 endemic species span a wide range, from 0.12 million years to 14.6 million years (Fig. 3). However, 76% of these

indicated with red markers, ten additional sampling sites with blue markers. Not indicated separately are 15 sites along the summit trail, and four sites very close to Mahua, Gunung Alab, and Inobong.

fall within the past 6 million years, the time span during which Kinabalu is likely to have reached its present elevation. Only two endemics, the frog *Kalophrynus baluensis* and the flowering plant *Ilex kinabaluensis*, are markedly older than the mountain itself. These may be explained as artefactual if we failed to identify the closest non-endemic sister lineage, thereby inflating their reconstructed age, or if these species are actually not endemics, but more widespread. Alternatively, they may truly be old endemics that evolved during cooler periods at lower elevations in Borneo before Kinabalu’s formation.

Our phylogenetic and biogeographic analyses (Extended Data Table 1 and Extended Data Figs 1–4, 7–9) suggest the existence of two categories of endemics¹⁸: ‘eccentric’ (12 taxa) and ‘centric’ (25 taxa). The eccentric type of endemic has sister taxa that occur either in temperate climates (seven cases) or in other tropical mountains outside of Borneo (five cases). To this group belong all bryophytes, pteridophytes, some of the fungal lineages and also the endemics in the flowering plant groups *Hedyotis*, *Ilex*, *Impatiens*, *Ranunculus* and *Euphrasia*, and the animals *Coelocia* and *Tritetrabdella*. Eccentric endemics predominantly occur at

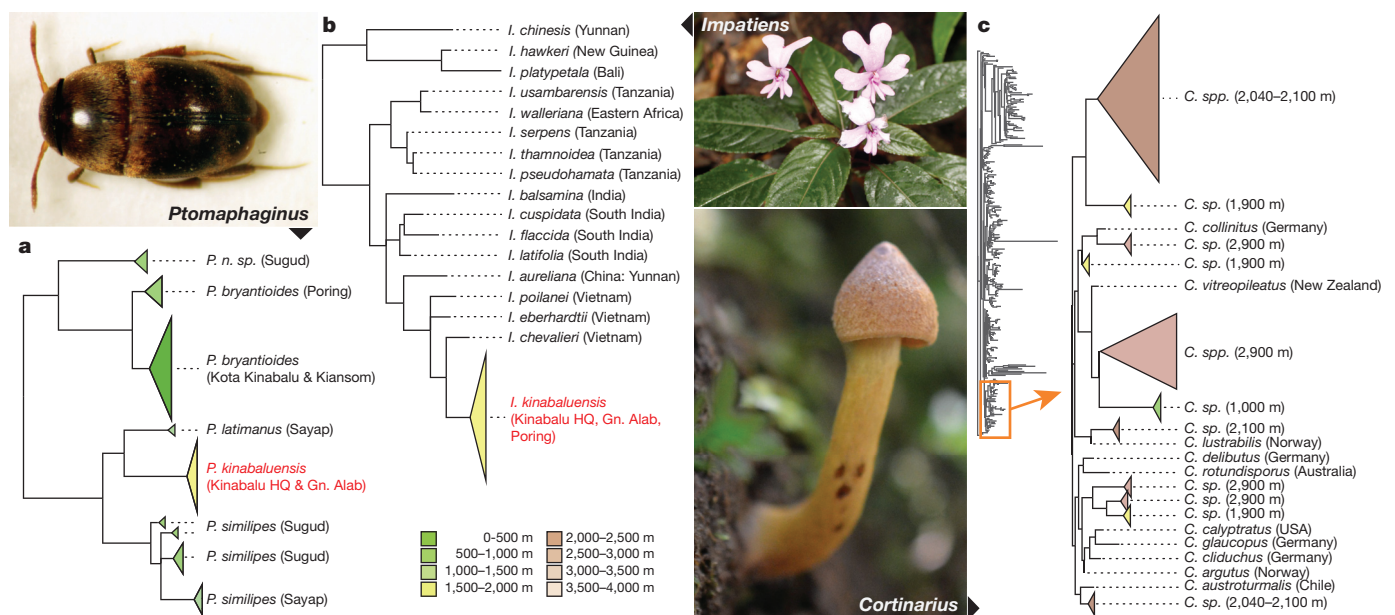


Figure 2 | Examples of phylogenetic reconstructions. a, Centric endemism (*Ptomaphagus*). b, Eccentric endemism (*Impatiens*). c, Centric endemism superimposed on eccentric endemism (detail of the *Cortinarius* phylogeny). Clades are coloured with the same colour scheme as in Fig. 1, by elevational

zone at which they occur (see legend). High-elevation endemics in a and b are indicated in red. For support values and other details on these and other trees, see Extended Data Figs 1–4.

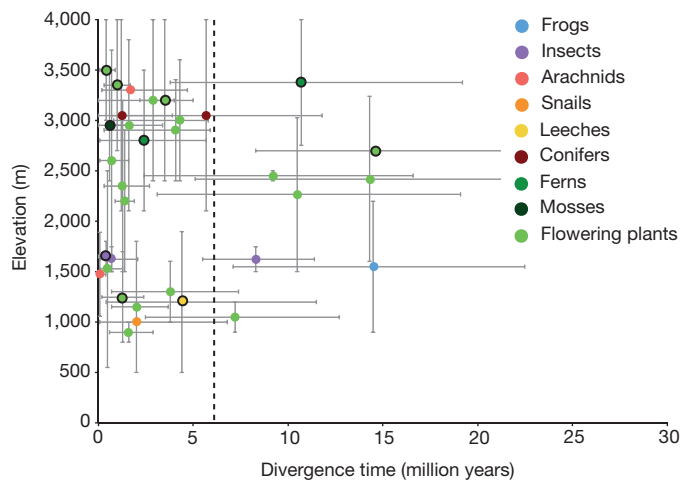


Figure 3 | Elevations and ages for endemic species. Elevations (mid-points, minima and maxima) and dates of origination derived from molecular dating (averages and 95% credible intervals) for endemic species; eccentric species (see main text) are indicated with a black circle. The vertical dashed line indicates the oldest possible date for Kinabalu to have reached its current elevation. For details, see Extended Data Figs 5 and 6.

high elevations (mean lower elevational boundary, 2,212 m; s.d., 837 m), they are strict Kinabalu endemics (they do not occur on nearby, lower mountains), and are further characterized by high dispersal capacities (one, two and seven clades with eccentric endemism have small, medium and large dispersal, respectively).

The centric type of endemic has sister taxa that occur locally, in the Borneo lowlands. Six out of the eight endemic animal species, some of the fungal lineages, the conifers, and 17 of the 20 flowering plants belong to this type. Centric endemics occur on average at lower elevations (mean lower elevational boundary, 1,724 m; s.d., 728 m) and tend to have lower dispersal capacities than the eccentric endemics (eight, seven and ten clades with centric endemism have small, medium and large dispersal, respectively). Some centric endemics are not strict Kinabalu endemics, as they are also found on other mountains in Kinabalu's vicinity. Our ancestral state reconstructions (Supplementary Table 6) confirm a pattern of altitudinal bottom-up cladogenesis in the centric endemics, with 8 out of 11 species ranging partially or entirely outside of the 95% credibility interval for the reconstructed elevation of the most recent common ancestor, and the remaining 3 out of 11 falling entirely within this interval. When we performed the same analysis with Kitayama's (ref. 19) seven vegetation zones, rather than elevation, we found that only three of these eight elevation shifts may represent a shift towards a higher vegetation zone. This suggests that, even in high-elevation centric endemics, substantial niche shifts are rare. Niche conservatism is probably even underestimated, since the Massenerhebung effect causes an upward shift of vegetation zones on Kinabalu compared with other Bornean mountains^{14,20}.

Our multi-taxon study shows that Kinabalu's biodiversity hot spot is of recent origin. This means the mountain is an evolutionary cradle, accumulating neo-endemics, as has been suggested for other young, high-elevation biodiversity hot spots at low latitudes²¹, such as the Tibetan plateau²², the Andean highlands⁸, and Afrotropical volcanoes²³. However, probably as a consequence of the rapid emergence of the mountain¹⁵ and its unique alpine summit conditions¹⁹, many of these neo-endemics have not evolved by drastic niche shifts from local ancestors, but rather by immigration of pre-adapted propagules from elsewhere. This explains the multiple independent colonization events in some taxa (for example, *Glomus*, *Rhododendron* and *Coeliccia*). In addition, local lowland taxa have also generated montane species, but although some of these have reached vegetation zones above 2,000 m, most do not show substantial niche shifts away

from their ancestral niche. The fact that the endemic biota of Kinabalu appears to be composed largely of pre-adapted (eccentric) species and locally derived (centric), ecologically conserved endemics is in line with niche conservatism²⁴.

We suggest that our novel approach of molecular dating of multiple clades be applied to larger communities in this and other tropical montane biodiversity hot spots²⁵. In combination, such information should allow a detailed dissection of the relative roles of ecological speciation, colonization and habitat filtering in the formation of endemic biotas in this and other tropical mountains. Moreover, such understanding could improve predictions of the likelihood of extinction and evolutionary rescue of endemic species experiencing changing climate conditions⁷.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 December 2014; accepted 14 July 2015.

Published online 12 August 2015.

- Körner, C. & Spehn, E. M. *Mountain Biodiversity: a Global Assessment* (Parthenon, Boca Raton, 2002).
- Chen, I. C. et al. Elevation increases in moth assemblages over 42 years on a tropical mountain. *Proc. Natl Acad. Sci. USA* **106**, 1479–1483 (2009).
- La Sorte, F. A. & Jetz, W. Projected range contractions of montane biodiversity under global warming. *Proc. R. Soc. Lond. B* **277**, 3401–3410 (2010).
- Graham, C. H. et al. The origin and maintenance of montane diversity: integrating evolutionary and ecological processes. *Ecography* **37**, 711–719 (2014).
- Cadena, C. D. et al. Latitude, elevational climatic zonation and speciation in New World vertebrates. *Proc. R. Soc. Lond. B* **279**, 194–201 (2012).
- Rodríguez-Castañeda, G. et al. Tropical forests are not flat: how mountains affect herbivore diversity. *Ecol. Lett.* **13**, 1348–1357 (2010).
- Schiffers, K., Borne, E. C., Lavergne, S., Thuiller, W. & Travis, J. M. J. Limited evolutionary rescue of locally adapted populations facing climate change. *Phil. Trans. R. Soc. Lond. B* **368**, 20120083 (2013).
- Hoorn, C., Mosbrugger, V., Mulch, A. & Antonelli, A. Biodiversity from mountain building. *Nat. Geosci.* **6**, 154 (2013).
- Warren, B. H. et al. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur–Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
- Watson, D. M. A conceptual framework for studying species composition in fragments, islands and other patchy ecosystems. *J. Biogeogr.* **29**, 823–834 (2002).
- Ornelas, J. F. et al. Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of northern Mesoamerica. *PLoS ONE* **8**, e56283 (2013).
- Madriñán, S., Cortés, A. J. & Richardson, J. E. Páramo is the world's fastest evolving and coolest biodiversity hotspot. *Front. Genet.* **4**, 192 (2013).
- Barkman, T. J. & Simpson, B. B. Origin of high-elevation *Dendrochilum* species (Orchidaceae) endemic to Mount Kinabalu, Sabah, Malaysia. *Syst. Bot.* **26**, 658–669 (2001).
- Gawin, D. F. et al. Patterns of avian diversification in Borneo: the case of the endemic Mountain Black-eye (*Chlorochrys emiliae*). *Auk* **131**, 86–99 (2014).
- Cottam, M. A. et al. Neogene rock uplift and erosion in northern Borneo: evidence from the Kinabalu granite, Mount Kinabalu. *J. Geol. Soc. Lond.* **170**, 805–816 (2013).
- Whitehead, J. *Exploration of Mount Kina Balu, North Borneo* (Gurney & Jackson, London, 1893).
- Wong, K. M. & Phillips, A. *Kinabalu, Summit of Borneo* (Sabah Society & Sabah Parks, Kota Kinabalu, 1996).
- van Steenis, C. G. G. J. Plant geography of the mountain flora of Mt. Kinabalu. *Proc. R. Soc. Lond. B* **161**, 7–38 (1964).
- Kitayama, K. An altitudinal transect study of the vegetation on Mount Kinabalu, Borneo. *Vegetatio* **102**, 149–171 (1992).
- Grubb, P. J. Interpretation of Massenerhebung effect on tropical mountains. *Nature* **229**, 44–45 (1971).
- Favre, A. et al. The role of the uplift of the Qinghai–Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev. Camb. Philos. Soc.* **90**, 236–253 (2015).
- Schwery, O. et al. As old as the mountains: the radiations of the Ericaceae. *New Phytol.* **207**, 355–367 (2015).
- Price, T. D. et al. Niche filling slows the diversification of Himalayan songbirds. *Nature* **509**, 222–225 (2014).
- Crisp, M. D. et al. Phylogenetic biome conservatism on a global scale. *Nature* **458**, 754–756 (2009).
- Culmsee, H. & Leuschner, C. Consistent patterns of elevational change in tree taxonomic and phylogenetic diversity across Malesian mountain forests. *J. Biogeogr.* **40**, 1997–2010 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the mountain guides of Sabah Parks for their help with material collection, ongoing assistance to continuing research in the area, and recent efforts following the 5 June 2015 earthquake. R. Butôt, M. de Jong, and

B. Driessen carried out DNA extractions and PCRs. P. Wilkie provided a tissue sample for *Payena*. For the fungi, T. Semenova extracted DNA from the soil samples and E. Duijm generated the reads on the Ion Torrent sequencer. B. Blankevoort and E. J. Bosch prepared Fig. 1. J. Holloway (Natural History Museum, London, UK), I. Das (Universiti Malaysia Sarawak), and T. S. Liew (Naturalis, UMS) provided background on Kinabalu endemism, and R. Hall (University of London) on its geological history. R. Etienne (University of Groningen), R. Hall (University of London) and F. Bossuyt (Free University Brussels) read a draft of the paper. Collection and export permits were supplied by the Sabah Biodiversity Centre. Besides institutional support via Netherlands FES-funding, this project received funding via grants 821.01.002 and 825.12.007 of the Netherlands Organization for Scientific Research, the Alberta Mennega Foundation, the Ecology Fund of the Royal Netherlands Academy of Sciences, the Uyttenboogaart-Eliassen Foundation, and Pro Acarologia Basiliensis. G. van Uffelen, R. van Vugt, and P. Kessler (Botanic Gardens, Leiden) and T. de Winter, G. Thijssen, B. Hoeksema, H. Dautzenberg, J. Mols, and E. van Nieuwerkerken (Naturalis) gave advice on transportation of specimens. Artist I. Kopelman and Naturalis general director E. van Huis (Naturalis) accompanied the expedition participants in the field. J. Gunsalam and P. Jilan (Sabah Parks Board of Trustees) officiated the opening and closing ceremony, respectively. P. Basintal, E. van Huis and E. Smets supervised the expedition as directors for the organizing institutions. Media coverage and communications were facilitated by J. van Alphen, A. Mohd. R. P. Dahlan, K. de Greef, A. Kromhout, J. Schilthuisen, E. E. Soudi, L. Stek, A. Taminiau, and M. Z. Abdul Wahab. Sabah Parks staff members L. Apin, B. Busin, B. Butit, J. Gampoyo, G. Gangko, B. Gangku, S. Gangku, P. Gimpiton, M. Gunggutou, T. Gunggutou, G. Gunsalam, J. Guntabid, I. Isnain, K. Ag. Jalil, Y. Johalin, R. M. Karim, Y. Kumin, D. Laban, A. Lahi, J. Lapidin, M. Latim, N. Maidin, N. Majuakim, R. Miadin, Y. Miki, M. Moguring, H. Mujih, F. Simin, K. C. Sion, D. Sumbin, D. Tingoh, E. Yakin, were invaluable for all manner of logistics and other support before, during, and after the expedition, in the field and at the stations and administrative offices. For

the Forest Research Centre researchers, J. B. Dawat, M. Gumbilil, A. J. Jaimin, J. Jumian, Jem. Jumian, J. L. Yukang, P. J. Miun, M. A. T. Mustapha, and B. Saludin assisted in the field. T. Siew of Borneo Indah Sdn. Bhd. was the local supplier for consumables and field equipment. We thank the ESEB Outreach Fund and the Treub Foundation for funding outreach activities in Kinabalu Park and the Crocker Range Park. N. Raes helped with the niche modelling.

Author Contributions M. Sch., V.S.F.T.M., M.Sp., M.M.B., C.B.M., L.E.B., K.T.C.A.P., J.N., R.R., and M.L. conceived and guided the study. M.Sch. and V.S.F.T.M. wrote the paper. K.H., F.S. and K.B. administrated all samples and sequences. K.H. and L.E.B. were responsible for DNA sampling in the field. V.S.F.T.M., A.A., N.A., A.B., P.-P.C., A.Y.C.C., R.D., H.F., C.F.-v.S., J.G., R.G., P.H., P.I., I.I., M.J., H.K., E.K., P.K., F.L., R.J.M., L.N.M., N.N., J.T.P., H.R., S.S., A.S., R.M.S., P.-S.S., H.S., N.S., J.B.S., M.Sul., S.Sum., J.v.T., F.Y.Y.T., B.E.Y., C.B.M., R.R. and M.Sch. collected material in the field and identified it following current taxonomy. K.H., V.S.F.T.M., H.d.B., F.A.A.F., J.G., B.G., P.H., S.B.J., S.N., M.St., D.C.T. and M.Sch. conducted the analyses.

Author Information GenBank accession numbers for newly determined sequences are: KF369344, KF492848, KF839944–KF839949, KM378672–KM378719, KM489525–KM489544, KM982697–KM982715, KP152323–KP152394, KP978313–KP979097, KR002164–KR002181, KR030167–KR030238 and KR232819–KR232836. A large portion of the data is also publicly available on BOLD (www.boldsystems.org), as dataset DS-KINA (<http://dx.doi.org/10.5883/DS-KINA>). Voucher specimens have been archived in public natural history collections. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.Sch. (Menno.Schilthuisen@naturalis.nl).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Expedition. Material was sampled chiefly between September 10th and 25th, 2012, by the participants of the Crocker Range/Kinabalu Scientific Expedition, which was conceived, organized, funded and conducted jointly by Sabah Parks (Malaysia) and Naturalis Biodiversity Center (The Netherlands). The expedition was set up with the explicit aim to investigate evolutionary patterns of endemism (see main text), and participants were selected to maximize taxon coverage and ecological diversity. The 47 expedition members were asked to identify one or more congeneric species complexes (hereafter termed 'genera') within their taxon of expertise that included high elevation endemics and preferably also more widespread lowland species (although there is a possibility that this approach would leave out any monotypic endemics, in the groups studied by us, such species are not known). These, plus outgroup species, were then sampled (using taxon-specific methods; see Supplementary Table 2 for further details) at eight expedition stations located at a range of elevations (Fig. 1), and also in additional locations in both parks and their vicinity. Data on time, latitude and longitude, elevation, and habitat were collected with each sample. Material was subsampled for DNA analysis, mostly in the field, by preserving tissues or entire organisms, usually either in pure ethanol or in silica gel (see Supplementary Table 2 for taxon-specific subsampling and storage). In the case of fungi, mycelial DNA was obtained by bulk extraction from composite soil cores (see Supplementary Table 2 and below). In addition to the data collected during the expedition, the expedition members were asked to provide a complete assessment of elevational and geographical distribution for each of the species within their taxon of expertise, based on published sources and previous collections.

Molecular analysis. Upon arrival in the Naturalis DNA barcoding facility, subsamples were processed in a semi-automated way. DNA extractions on animal material were performed on either tissue samples or individual arthropod legs, or using whole arthropods in a non-destructive extraction method. DNA was extracted using the NucleoMag 96 Tissue kit (Macherey-Nagel GmbH & Co., Düren, Germany) on the KingFisher Flex magnetic particle processor (Thermo Scientific). DNA extractions on plant tissue were performed on dried leaf samples, using the NucleoMag 96 Plant kit (Macherey-Nagel GmbH & Co., Düren, Germany). The KingFisher Flex protocols were slightly adjusted from the manufacturer's protocol (<http://www.thermoscientific.com>) to include longer mixing of beads with lysed sample and a drying step before elution. A volume of 150 µl was used for elution in both animal and plant samples. Soil samples containing fungal mycelia were processed by pooling 40 cores (for sampling details, see Supplementary Table 2), resulting in a composite soil sample for each site. Genomic DNA was extracted from 1 g of this sample using the NucleoSpin Soil kit (Macherey-Nagel GmbH & Co., Düren, Germany), according to the manufacturer's protocol. For each sample DNA extraction was carried out twice and replicates were combined.

For each specimen, the accepted barcode marker for its taxon was sequenced, in many cases augmented with one or more other phylogenetically informative loci. Supplementary Tables 1 and 2 provide lists of the loci and primers used for each genus included in the study. Where relevant for outgroup comparison or further geographic resolution, we then obtained additional sequences from Borneo or elsewhere from GenBank or from our own, unrelated studies. GenBank accession codes for these are given in Supplementary Table 2. After sequencing, (remains of) voucher specimens were mostly deposited in the collections of Sabah Parks and/or Naturalis Biodiversity Center. Voucher collection codes are listed in Supplementary Table 2.

PCR reactions generally contained 18.75 µl ultrapure H₂O, 2.5 µl 10× PCR buffer, 1.0 µl 10 mM of each primer, 0.5 µl 2.5 mM dNTPs and 0.25 µl 5U µl⁻¹ Qiagen Taq. Most PCRs were performed using an initial denaturation step of 180 s at 94 °C, followed by 40 cycles of 15 s at 94 °C, 30 s at 50 °C and 40 s at 72 °C, and ended with a final extension of 300 s at 72 °C and pause at 12 °C. Bi-directional Sanger sequencing was performed at either Macrogen (<http://www.macrogen.com>) or BaseClear (<http://www.baseclear.com>). Sequences were edited manually (primers trimmed off and obvious reading errors corrected) using Sequencher 4.10.1 (Gene Codes Corporation). For the fungal soil samples, the amplicon library was sequenced using an Ion 318™ Chip in an Ion Torrent Personal Genome Machine (PGM; Life Technologies, Guilford, CT, USA). PCR, sequencing, raw data filtering and all bioinformatic analyses were performed as described in detail previously²⁶. From these fungal soil sequences, we selected non-redundant sequences of the ectomycorrhizal genus *Cortinari* and the arbuscular mycorrhizal genus *Glomus* because they were distributed along the entire altitudinal range of our sampling.

Phylogenetic analysis. We performed initial exploratory alignment and created neighbour-joining trees. This allowed an assessment of individuals that (on the basis of cursory morphological examination) had been included as potential (undescribed) members of the species complexes of choice. In some cases, the sequence analysis showed that these individuals were in fact unrelated and such sequences were then removed before further analysis. Then, using Geneious (version 6.0.5; created by Biomatters; available from <http://www.geneious.com/>) as a workbench, sequence data were aligned with MUSCLE²⁵. We ran ModelTest²⁷ to determine, using Akaike Information Criterion (AIC), the best model of nucleotide substitution per alignment, for each genetic marker separately. We then performed a Bayesian analysis using MrBayes 3.2.2 (ref. 28) with settings for multiple markers unlinked, running two replicates of a minimum of 10⁷ generations each. Convergence was checked in Tracer (v1.5; ref. 29) and, after discarding a 25% burn-in, the resulting majority-rule consensus trees were preserved. For the fungal data sets, multiple sequence alignments were built using MUSCLE and, due to the large size of these data sets, phylogenetic analyses of maximum likelihood (ML) were carried out in Garli 2.0 (ref. 30) using the GTR+I+G model with estimated values of the alpha parameter of the gamma distribution of variable sites (G) and proportion of invariable sites (I). Clade support was assessed by non-parametric bootstrap analyses with 100 pseudo-replicates.

Molecular dating. The times of divergence of Mount Kinabalu endemics from their most closely related extant relatives were estimated with a molecular clock approach. We included not only our own data, but also incorporated published studies on genera other than the ones we sampled. For each genus, taxon sampling and molecular data used are specified in Supplementary Table 4. All data sets were assembled in Geneious (ver. 5.5.6; created by Biomatters; available from <http://www.geneious.com/>). Data sets were pruned to contain a single taxon per species by selecting a random sequence from each species. Sequence alignments were obtained with the MAFFT (ver. 6.814b) alignment tool implemented in Geneious. Divergence time estimates were obtained using a relaxed molecular clock approach with BEAST (ver. 2.1.3; ref. 31). For each DNA region used, an unlinked substitution model was implemented as selected with jModeltest (ver. 2.1.5; ref. 27) under the AIC. For the tree prior, a Yule birth rate prior was selected with a flat prior on the Yule birth rate. A single clock and tree model was used for each analysis. All analyses were run with an uncorrelated log-normal clock model, to which an exponential prior was assigned for both the mean (ucldMEAN; mean set at 10.0) and the standard deviation (ucldStdev; mean set at 0.33). Time calibration was done using: (1) fossil constraints; (2) secondary calibration points; or (3) substitution rates, depending on the information available for each particular genus. In cases where suitable fossil or secondary calibration information was not available, we used a priori information about the substitution rate of one of the gene partitions—as opposed to the 'fix the mean substitution rate' option across all partitions—to optimize the analyses. The full details on calibration methods used for each taxon are given in Supplementary Table 5. For each run, posterior distributions of parameters were approximated using a Markov chain Monte Carlo analysis of 5 × 10⁷ generations, sampling every 5,000 generations, discarding a burn-in of 5 × 10⁶ generations (10%). Convergence of the chains was checked by evaluating the ESS values of each parameter with Tracer v1.6.

Biogeographical, elevational and vegetation zone analyses. For each genus, an independent estimation of the dispersal neighbourhood size was obtained from the spatial ecology literature on the genus concerned or related genera, and/or expert knowledge. We used this information to classify dispersal neighbourhood (two standard deviations for the distribution of distances between birth and reproduction for individuals of a certain species) in the genus in question as 'small', <1 km (mostly flightless animals), 'medium', 1–100 km (flying or ballooning animals, and plants dispersed by flying animals), or 'large', >100 km (plants and fungi with very small wind-dispersed seeds or spores). An overview of dispersal assessments is in Supplementary Table 3.

Ancestral ranges were estimated for each taxon group under maximum likelihood with BioGeoBears³² using the three-parameter DEC+J model. The ultrametric tree from the Bayesian relaxed molecular clock analysis was used as input tree. This tree was pruned to include the clade with the endemic species and at least one deeper node (for most groups two deeper nodes were included). Each species was scored to be absent or present in Mount Kinabalu (K), Borneo (B), and the rest of the world (W). For each group the analysis was run with equal probability for exchange between the three areas over the entire diversification time of the tree. The maximum number of areas was assigned as the maximum number of areas observed in the extant taxa (either 2 or 3). The optimizations, with the ancestral range with the highest likelihood for each node, are shown in Extended Data Figs 7–9). We classify a 'W' to 'K' or 'KW' shift in the lineage leading to a species endemic to Kinabalu as 'eccentric endemism'. All other ancestral ranges leading to Kinabalu endemics are scored as 'centric'.

For each genus sampled during the expedition that showed centric endemism, we added the elevation at which it had been collected as a continuous character. In BayesTraits 2.0 (ref. 33) we reconstructed ancestral states for elevation at the node leading to the endemic species and its sister clade using the continuous random walk model which implements a generalized least squares model for continuously varying traits. The elevation of the most-recent common ancestor of the node of interest was inferred using a MCMC approach on a set of 2,700 trees randomly sampled after the phylogenetic inference with MrBayes (excluding burn-in). The MCMC chain of BayesTraits was run for 10^6 generations, discarding the results sampled during the first 10^5 generations as burn-in. We then inspected whether the current range of elevation of the endemic species stretches beyond the 95% credible interval for the most-recent common ancestor. If it stretches above this interval, we considered this indicative of a 'bottom-up' scenario, whereas the reverse would suggest a 'top-down' scenario. The results are summarized in Supplementary Table 6.

In parallel with the elevational analysis, we conducted an analysis in which we coded the vegetation belt(s) as taxon traits. We used the zones defined for Mt. Kinabalu in ref. 19, namely: (i) lowland vegetation (below 600 m); (ii) hill vegetation (600–1,200 m); (iii) lower montane vegetation (1,200–2,000 m); (iv) upper montane vegetation (2,000–2,800 m); (v) subalpine vegetation (2,800–3,400 m); (vi) upper subalpine vegetation (3,400–3,700 m); (vii) alpine vegetation (above 3,700 m). The probabilities at each vegetation belt for the most-recent common ancestor of the node leading to the endemic clade and its sister group were inferred with the continuous-time Markov model for discrete data implemented in BayesTraits 2.0. The same input trees were used as for the elevation reconstruction, and the MCMC was run for 10^6 generations, discarding the results sampled during the first 10^5 generations as burn-in. The resulting probabilities are shown in Supplementary Table 7.

Geological summary. The following is based on refs 15,34 and R. Hall, personal communication, 2014.

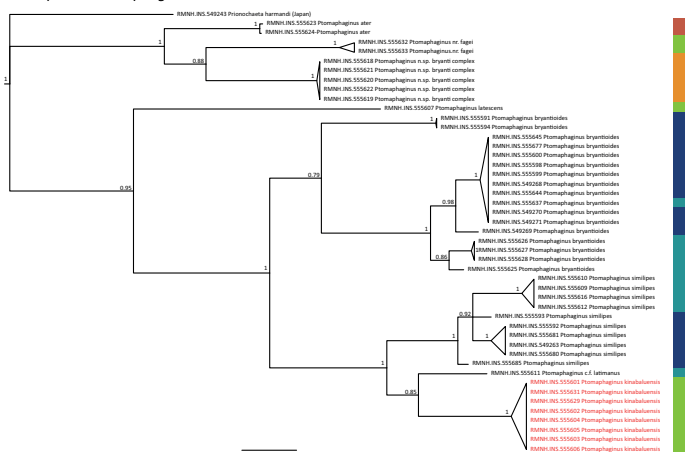
Orogeny in the region began at about 23 million years ago (Mya) as a result of the northward movement of the Australian plate leading to the collision of Australia and Southeast Asia in what today is eastern Indonesia. Orogeny in the Kinabalu area began at about 20 Mya after collision of north Borneo with the South China continental crust. This led to the emergence of most of Sabah. Soon

after this emergence much of Sabah became submerged, although a north–south-oriented mountain belt formed of tightly folded mudstones and sandstones (the future Crocker Range) remained. At the same time, the climate changed from seasonal to ever-wet. Mapping of former shelf edges offshore shows that later the Crocker Range widened with time, implying that it became higher in its central parts. Between 8 and 7 Mya a granitic body intruded the crust and cooled rapidly from ~7 to 5 Mya. Melting could have resulted from a build-up of heat from radioactive decay in the deep crust, or from decompression melting following extension. After 6 Mya (the precise timing and duration is not known) the area (including the surrounding Crocker Range) rose further, elevating to the point that it was glaciated recently, possibly after a lithospheric root detached and sank into the deeper mantle, leading new mantle material to flow into the vacated space below the Crocker Range. Extensive glacial weathering of the summit suggests the mountain had reached its present elevation by at least 1 Mya. The Pinosuk gravels on Kinabalu's south-eastern slope are the accumulated erosional debris of young glaciation. Today, the mountain is still rising by 0.5 mm per year.

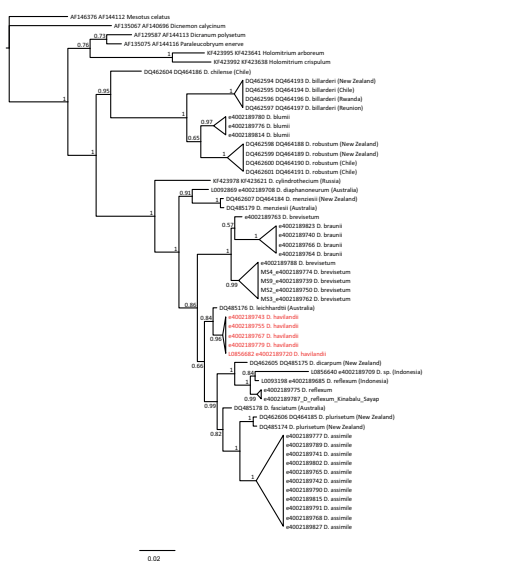
26. Geml, J. *et al.* Large-scale fungal diversity assessment in the Andean Yungas forests reveals strong community turnover among forest types along an altitudinal gradient. *Mol. Ecol.* **23**, 2452–2472 (2014).
27. Posada, D. & Crandall, K. A. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
28. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
29. Rambaut, A. & Drummond, A. J., *Tracer v1.5*. <http://beast.bio.ed.ac.uk/Tracer> (2009).
30. Zwickl, D. J. *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion* (The University of Texas, Austin, 2006).
31. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
32. Matzke, N. J. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **5**, 242–248 (2013).
33. Meade, A. & Pagel M. BayesTraits, version 2.0. <http://www.evolution.rdg.ac.uk> (2015).
34. Hall, R. *et al.* *The Geology of Mount Kinabalu* (Sabah Parks, 2008).



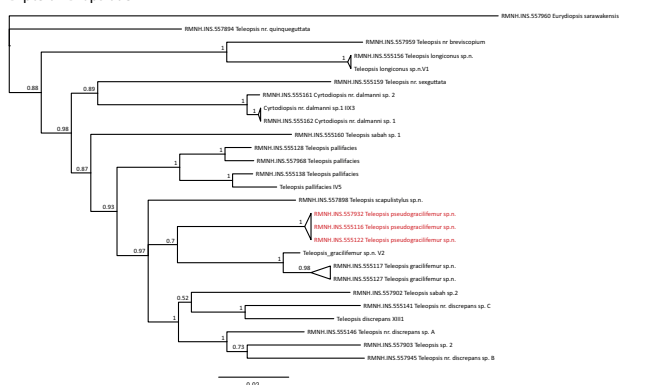
Coleoptera - Ptomaphagus



Dicranaceae - Dicranoloma



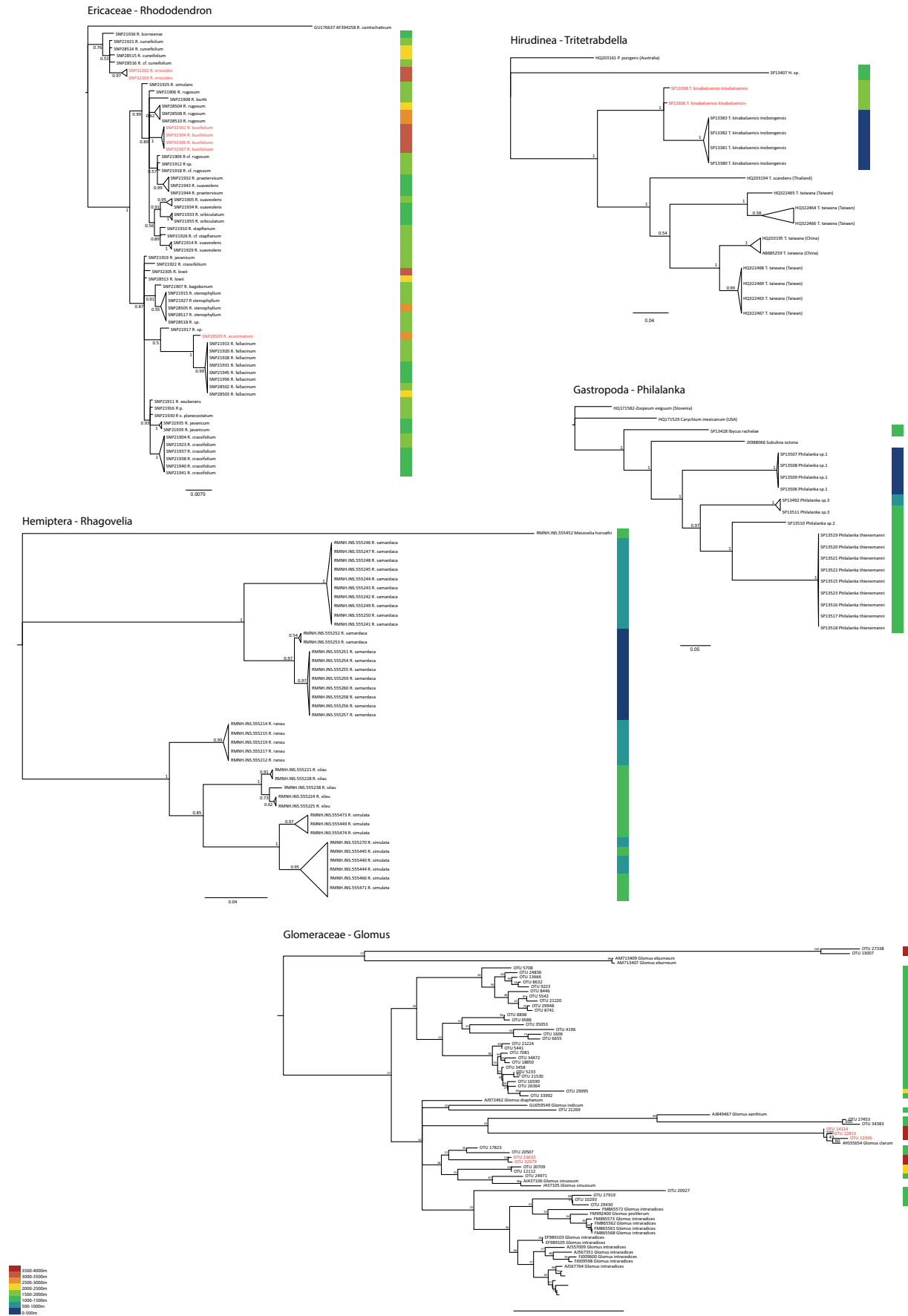
Diptera - Diopsidae



Cortinariaceae - Cortinarius



Extended Data Figure 2 | Phylogenetic reconstructions for Coleoptera, Dicranaceae, Diptera and Cortinariaceae. Details of materials and methods used are given in Supplementary Table 2.

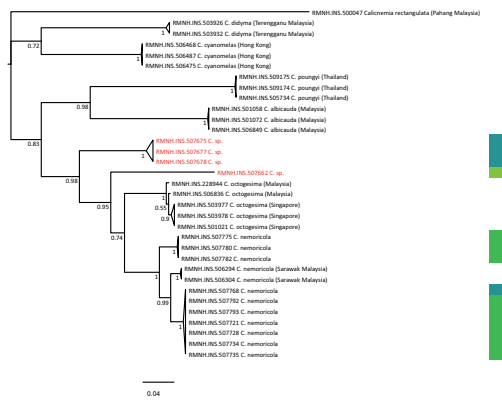


Extended Data Figure 3 | Phylogenetic reconstructions for Ericaceae, Gastropoda, Glomeraceae, Hemiptera and Hirudinea. Details of materials and methods used are given in Supplementary Table 2.

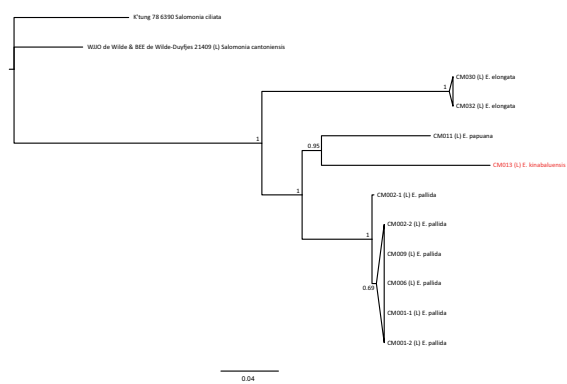
Nepenthaceae - Nepenthes



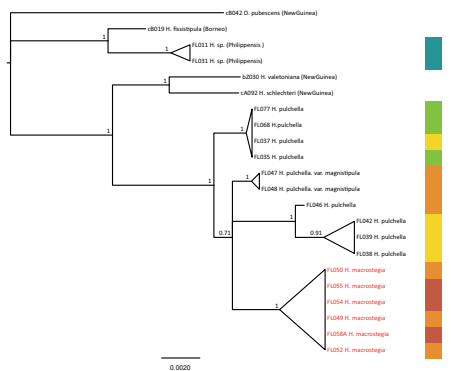
Odonata - Coeliccia



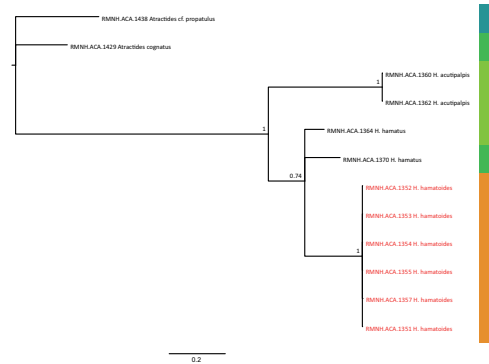
Polygalaceae - Epirixanthes



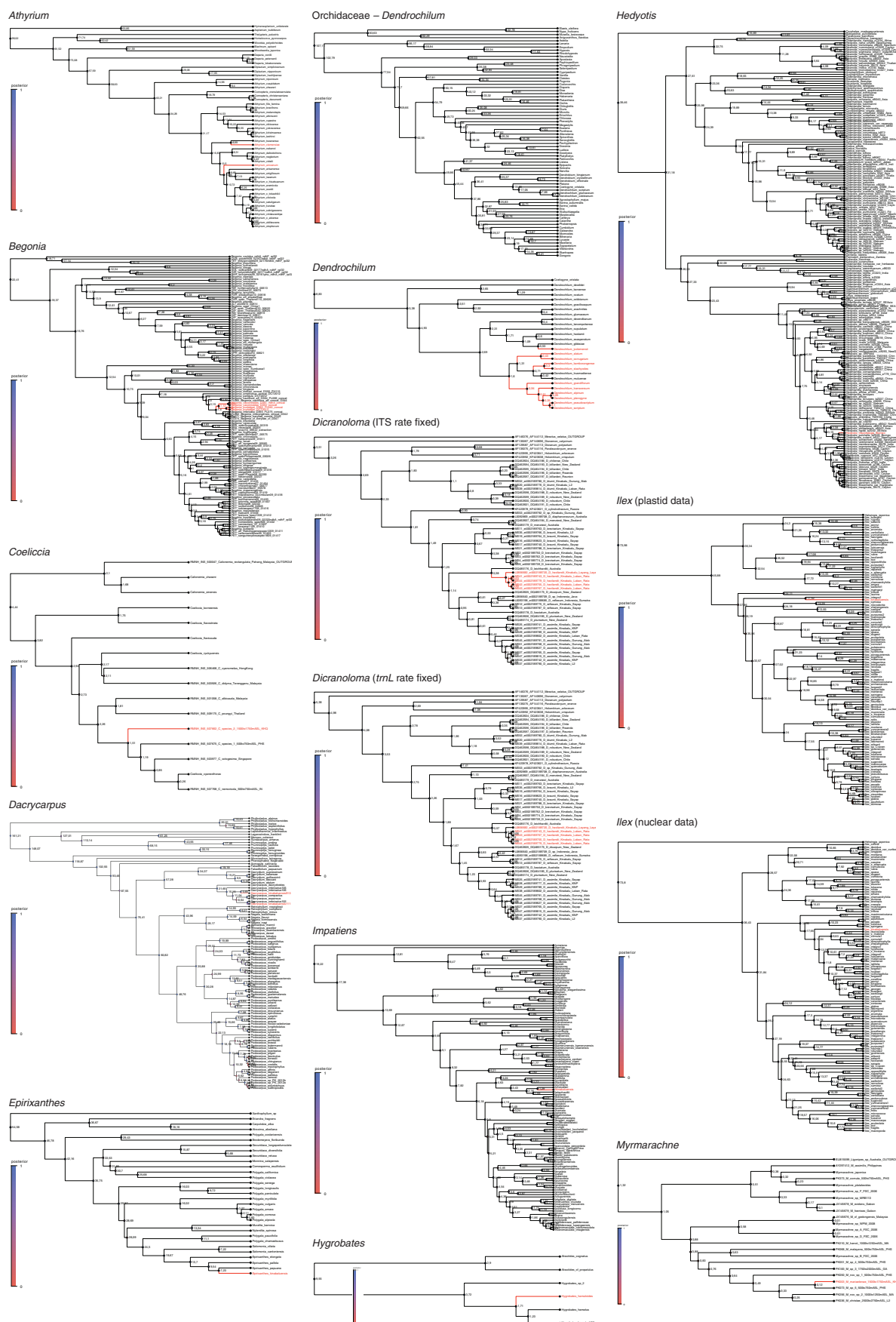
Rubiaceae - Hedyotis



Trombidiformes - Hygrobatas

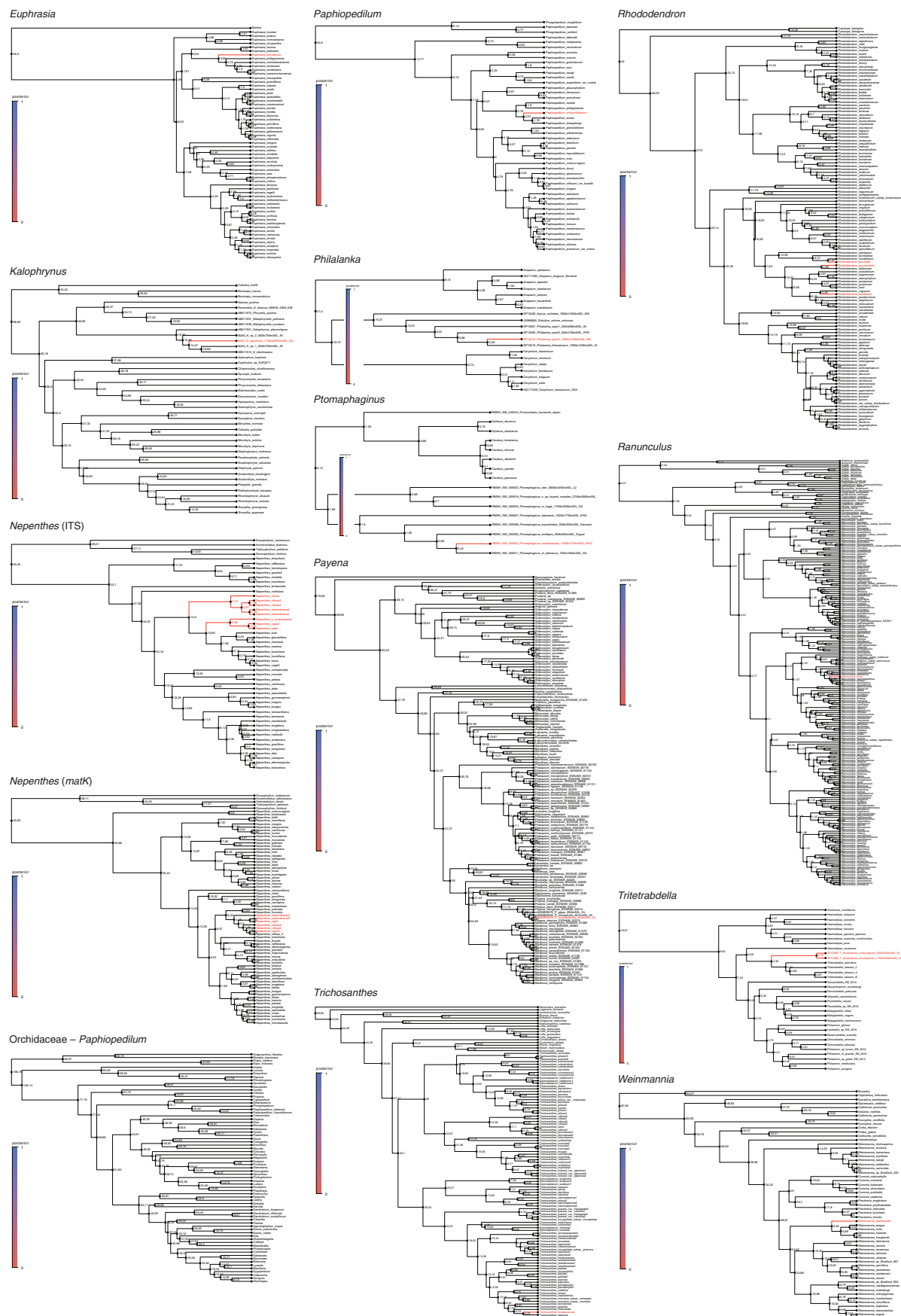


Extended Data Figure 4 | Phylogenetic reconstructions for Nepenthaceae, Odonata, Polygalaceae, Rubiaceae and Trombidiformes. Details of materials and methods used are given in Supplementary Table 2.



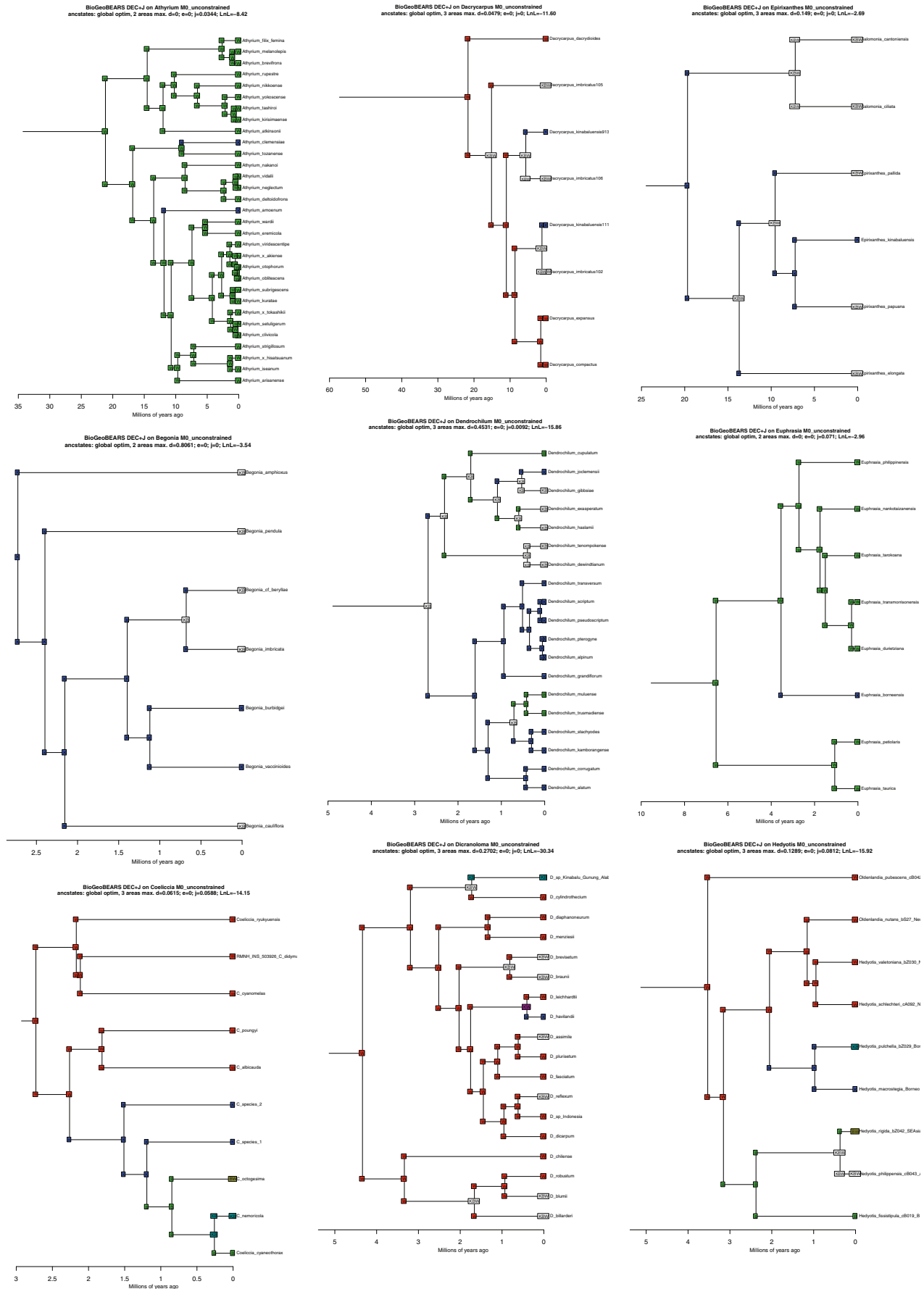
Extended Data Figure 5 | Chronograms for *Athyrium*, *Begonia*, *Coelliccia*, *Dacrycarpus*, *Epirixanthes*, *Dendrochilum*, *Dicranoloma*, *Impatiens*, *Hygrobatas*, *Hedyotis*, *Ilex* and *Myrmarachne*. See Supplementary Table 5

for more details. Node colours represent posterior probabilities. Endemic species printed in red.

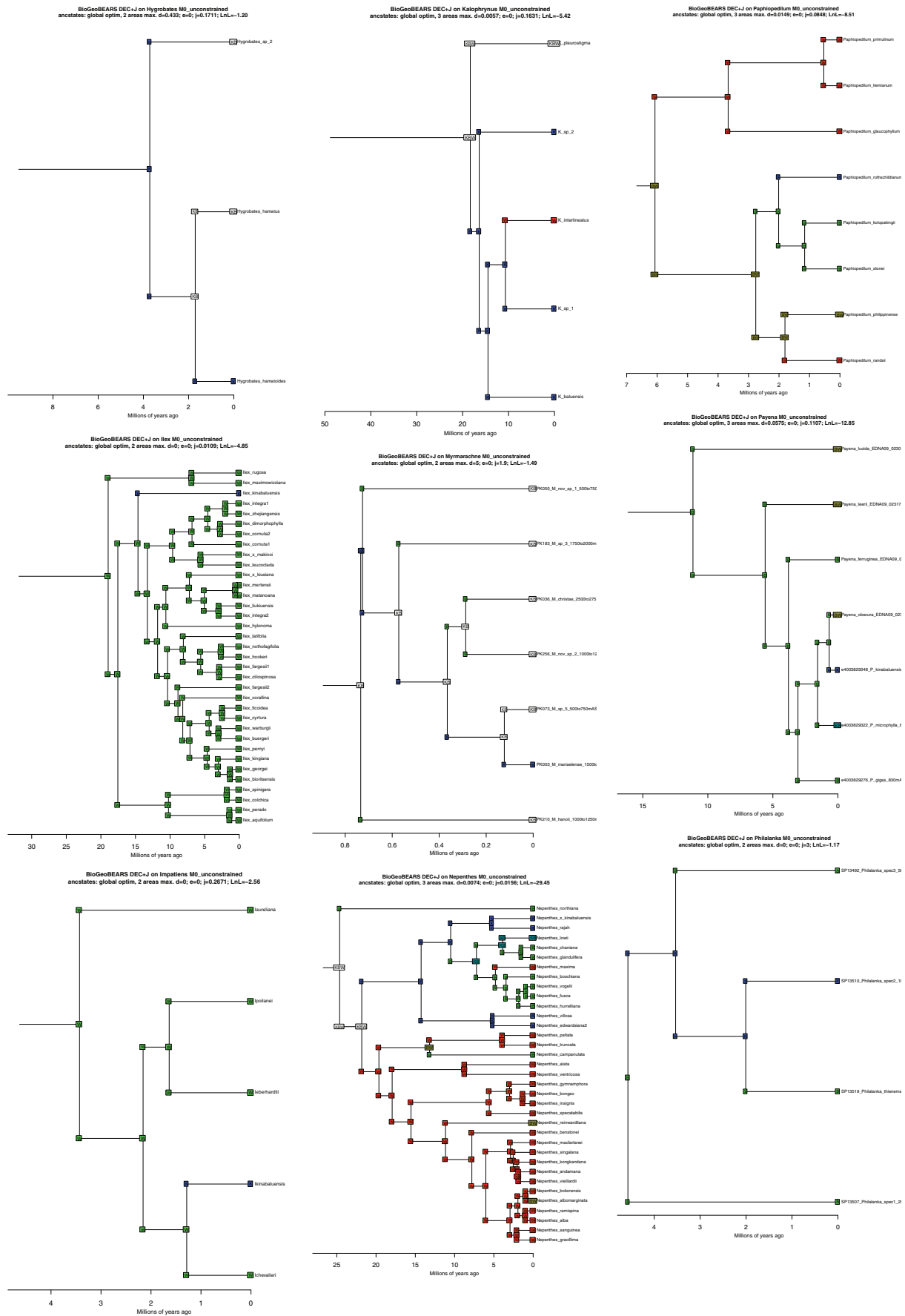


Extended Data Figure 6 | Chronograms for *Euphrasia*, *Kalophrinus*, *Nepenthes*, *Paphiopedilum*, *Philalanka*, *Ptomaphagus*, *Payena*, *Trichosanthes*, *Rhododendron*, *Ranunculus*, *Tritetrabdella* and

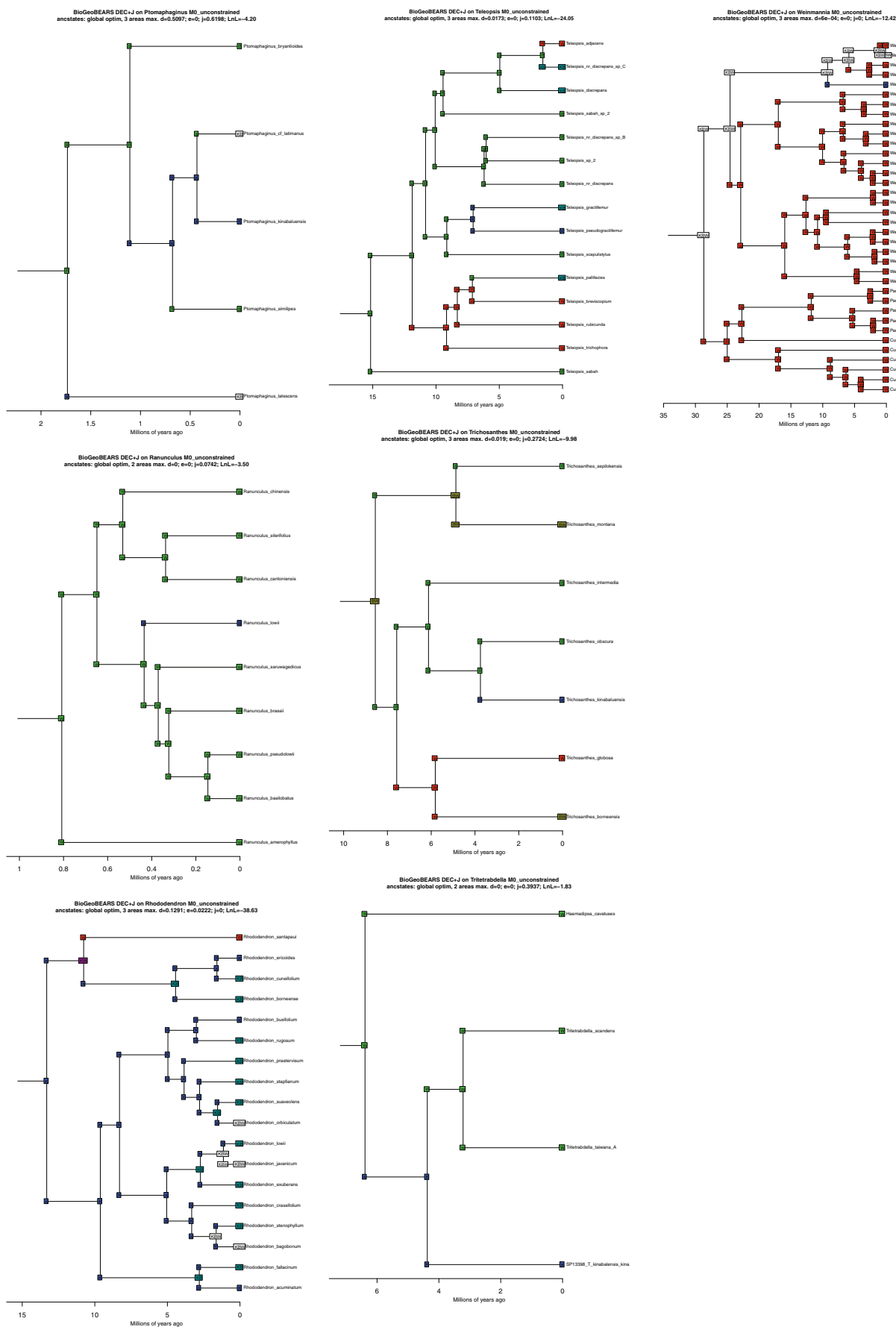
Weinmannia. See Supplementary Table 5 for more details. Node colours represent posterior probabilities. Endemic species printed in red.



Extended Data Figure 7 | Ancestral range estimations for *Athyrium*, *Begonia*, *Coeliccia*, *Dacrycarpus*, *Dendrochilum*, *Dicranoloma*, *Epirixanthes*, *Euphrasia* and *Hedyotis*. See Methods for further details. The trees are alphabetically ordered by genus name.



Extended Data Figure 8 | Ancestral range estimations for *Hygrobates*, *Illex*, *Impatiens*, *Kalophrynus*, *Myrmarachne*, *Nepenthes*, *Paphiopedilum*, *Paysonia* and *Philalanka*. See Methods for further details. The trees are alphabetically ordered by genus name.



Extended Data Figure 9 | Ancestral range estimations for *Ptomaphagus*, *Ranunculus*, *Rhododendron*, *Teleopsis*, *Trichosanthes*, *Tritetrabdella* and *Weinmannia*. See Methods for further details. The trees are alphabetically ordered by genus name.

Extended Data Table 1 | Details for all studied endemic taxa

Endemic taxon	Origin**	Dispersal capacity***	Region****	Elevation range (m)	Mean divergence (mya)	95% CI	Occurrence of sister lineage
<i>Dicranoloma havilandii</i>	eccentric	Large	K	2400 – 3500	0.4	0 – 1.3	Australia
<i>Athyrium clemensiae</i>	eccentric	Large	K	2750 – 4000	10.7	3.8 – 19.2	China, Japan (800–2600 m)
<i>Athyrium amoenum</i>	eccentric	Large	K	2100 – 3500	2.4	0.1 – 5.7	China, Japan (1500–3000 m)
<i>Dacrycarpus kinabaluensis</i> 1	centric	Small	K	2100 – 4000	5.7	0.7 – 11.8	Tropical SE Asia, incl. Kinabalu (700–2700 m)
<i>Dacrycarpus kinabaluensis</i> 2	centric	Small	K	2100 – 4000	1.2	0 – 3.9	Tropical SE Asia, incl. Kinabalu (700–2700 m)
<i>Nepenthes edwardsiana</i> - <i>N. vilosa</i>	centric	Small	K,T	1600 – 3240	14.3	5.1 – 25.9	Tropical SE Asia, incl. mountains on Borneo (600–2650 m)
<i>Nepenthes rajah</i> - <i>N. x kinabaluensis</i>	centric	Small	K,T	1500 – 3030	10.5	3.1 – 19.1	Tropical SE Asia, incl. mountains on Borneo (600–2600 m)
<i>Epirixanthes kinabaluensis</i>	centric	Small	K	900 – 1200	7.2	2.5 – 12.7	Tropical SE Asia, incl. Kinabalu
<i>Trichosanthes kinabaluensis</i>	centric	Large	K	1000 – 1600	3.8	0.7 – 7.4	Borneo lowlands
<i>Ilex kinabaluensis</i>	eccentric	Large	K	2700	14.6	8.3 – 21.7	East Asia
<i>Impatiens kinabaluensis</i>	eccentric	Large	K,C	800 – 1700	1.3	0.2 – 2.4	Vietnam
<i>Begonia vaccinoides</i> - <i>B. burbridgei</i>	centric	Small	K,T	1500 – 2900	1.4	0.9 – 1.9	North Borneo (incl. Kinabalu), montane forest
<i>Dendrochilum</i> clade 1	centric	Large	K	2100 – 3800	0.9	0.2 – 2.1	Borneo mountains (incl. Kinabalu)
<i>Dendrochilum</i> clade 2	centric	Large	K	1500 – 3200	1.3	0.3 – 2.7	Borneo mountains (incl. Kinabalu)
<i>Dendrochilum</i> clade 3	centric	Large	K	1500 – 3700	0.7	0.1 – 1.6	Borneo mountains (not Kinabalu)
<i>Dendrochilum</i> clade 4	centric	Large	K	1050 – 2000	0.5	0.1 – 1.3	Borneo mountains (incl. Kinabalu)
<i>Paphiopedilum rothschildianum</i>	centric	Large	K	500 – 1800	2.0	0.7 – 3.7	Borneo lowland
<i>Ranunculus lowii</i>	eccentric	Large	K	3000 – 4000	0.4	0.1 – 0.9	New Guinea
<i>Hedyotis macrostegia</i>	eccentric	Medium	K,T	2700 – 4000	1.0	0.3 – 1.7	Borneo (incl. Kinabalu)
<i>Euphrasia borneensis</i>	eccentric	Large	K	2400 – 4000	3.5	2.2 – 5.0	East Asia
<i>Paysona kinabaluensis</i>	centric	Medium	K	800 – 1000	1.6	0.6 – 2.9	Borneo (lowland and lower montane forest)
<i>Rhododendron ericoides</i>	centric	Medium	K	2400 – 4000	2.9	0.0 – 4.0	North Borneo (incl. Kinabalu), lower and mid montane forest
<i>Rhododendron buxifolium</i>	centric	Medium	K	2400 – 3600	4.3	0.8 – 5.8	North Borneo (incl. Kinabalu), montane forest
<i>Rhododendron acuminatum</i>	centric	Medium	K	2400 – 3400	4.1	0.3 – 5.9	East Asia (incl. Borneo), montane forest
<i>Weinmannia clemensiae</i>	centric	Large	K	2400 – 2500	9.2	3.4 – 16.6	Fiji and Malaysia (incl. Borneo), lowland to lower montane forest
<i>Glomus</i> clade 1	centric	Large	K	2850 – 3600	n/a	n/a	Kinabalu and Crocker Range, <2850 m elev.
<i>Glomus</i> clade 2	eccentric	Large	K	2850 – 3600	n/a	n/a	Europe
<i>Cortinarius</i> clades 1, 3, 4, 10, 11, 12, 14, 18, 20, 21, 23, 24, 26, 27, 28, 31*	eccentric	Large	K	2850 – 3600	n/a	n/a	Holarctic, South Pacific
<i>Cortinarius</i> clades 2, 5–9, 13, 15–17, 19, 22, 25, 29, 30*	centric	Large	K	2850 – 3600	n/a	n/a	Kinabalu and Crocker Range, <2850 m elev.
<i>Teleopsis pseudograssifemur</i> *****	centric	Medium	K, C	1500 – 1750	8.3	5.5 – 11.4	Kinabalu, Crocker Range, lower montane forest
<i>Ptomaphagus kinabaluensis</i>	centric	Medium	K, C	1500 – 1800	0.4	0.0 – 0.7	Kinabalu, Crocker Range, lowland and lower montane forest
<i>Tritetrabdella kinabaluensis</i>	eccentric	Small	K, C	500 – 1900	4.4	0.4 – 11.5	Malay Peninsula, Thailand, China, Taiwan, Japan, mountainous forests (undersampled)
<i>Philalanka</i> sp. 2	centric	Small	C, TM	1000	2.0	0.1 – 6.8	Southeast Asia (incl. Borneo), lowland and lower montane forest
<i>Coellicia</i> sp. 2	eccentric	Medium	K	1500 – 1750	0.7	0.1 – 2.1	Southeast Asia (incl. Borneo), lowland and lower montane forest
<i>Myrmarachne mariaelenae</i>	centric	Medium	K, C	1058 – 1888	0.1	0.0 – 0.4	Kinabalu, lowland forest
<i>Hygrobates hamatoides</i>	centric	Large	K	3300	1.7	0.2 – 4.7	Southeast Asia + Australia, incl. Borneo; in lowland and hill forest
<i>Kalophrynus baluensis</i>	centric	Small	K, C	1300 – 1800	14.5	7.1 – 22.5	Sundaland, incl. Borneo; lowland and hill forest

Plants are highlighted in green, fungi in blue, and animals in brown.

* For clade numbering in *Cortinarius*, see Supplementary Table 2.

** Centric/eccentric status determined via biogeographic analysis (see Extended Data 4).

*** For definitions and sources, see Supplementary Table 3.

**** K, Mount Kinabalu; T, Mount Tambuyukon; C, Crocker Range; TM, Mount Trus Madi.

***** Manuscript name.

A self-organized biomechanical network drives shape changes during tissue morphogenesis

Akankshi Munjal¹, Jean-Marc Philippe¹, Edwin Munro² & Thomas Lecuit¹

Tissue morphogenesis is orchestrated by cell shape changes. Forces required to power these changes are generated by non-muscle myosin II (MyoII) motor proteins pulling filamentous actin (F-actin). Actomyosin networks undergo cycles of assembly and disassembly (pulses)^{1,2} to cause cell deformations alternating with steps of stabilization to result in irreversible shape changes^{3–6}. Although this ratchet-like behaviour operates in a variety of contexts, the underlying mechanisms remain unclear. Here we investigate the role of MyoII regulation through the conserved Rho1–Rok pathway⁷ during *Drosophila melanogaster* germband extension. This morphogenetic process is powered by cell intercalation, which involves the shrinkage of junctions in the dorsal–ventral axis (vertical junctions) followed by junction extension in the anterior–posterior axis⁸. While polarized flows of medial–apical MyoII pulses deform vertical junctions, MyoII enrichment on these junctions (planar polarity) stabilizes them⁶. We identify two critical properties of MyoII dynamics that underlie stability and pulsatility: exchange kinetics governed by phosphorylation–dephosphorylation cycles of the MyoII regulatory light chain; and advection due to contraction of the motors on F-actin networks. Spatial control over MyoII exchange kinetics establishes two stable regimes of high and low dissociation rates, resulting in MyoII planar polarity. Pulsatility emerges at intermediate dissociation rates, enabling convergent advection of MyoII and its upstream regulators Rho1 GTP, Rok and MyoII phosphatase. Notably, pulsatility is not an outcome

of an upstream Rho1 pacemaker. Rather, it is a self-organized system that involves positive and negative biomechanical feedback between MyoII advection and dissociation rates.

To understand the mechanisms that underlie pulsatility (shape deforming) and stability (shape stabilizing) of actomyosin networks during morphogenesis, we investigated the role of the Rho1 pathway during germband extension (GBE)⁹. Rho1 GTPase activates Rho kinase (Rok)⁷, which in turn phosphorylates the MyoII regulatory light chain (RLC), resulting in assembly of processive bipolar filaments that bind F-actin to generate contractility¹⁰. Rok also inhibits the myosin-binding subunit (MBS)¹¹ of the MyoII phosphatase complex, responsible for dephosphorylating the RLC and thus inactivating the motor¹². Previous studies, with the use of phosphomimetic RLC (T20E/S21E)¹³, have reported that MyoII activation during gastrulation^{8,14,15} and planar polarity¹⁶ requires Rok-dependent phosphorylation. Phosphomimetic RLC, by definition locked in an active state, results in reduced GBE¹⁶, suggesting that dephosphorylation by phosphatase is essential. To test this, we inhibited the RLC-targeting subunit of the MyoII phosphatase complex (Mypt-75d (ref. 17)) using RNA interference (RNAi) (Extended Data Fig. 1a). RNAi inhibition of Mypt-75d resulted in slower GBE and junction shrinkage (Extended Data Fig. 1b, c). Green fluorescent protein (GFP)-tagged wild-type RLC revealed that the enrichment of MyoII on vertical versus transverse junctions (along the anterior–posterior axis) was 80% in controls, and reduced to 30% in Mypt-75d- and MBS-inhibited embryos (Fig. 1a, b

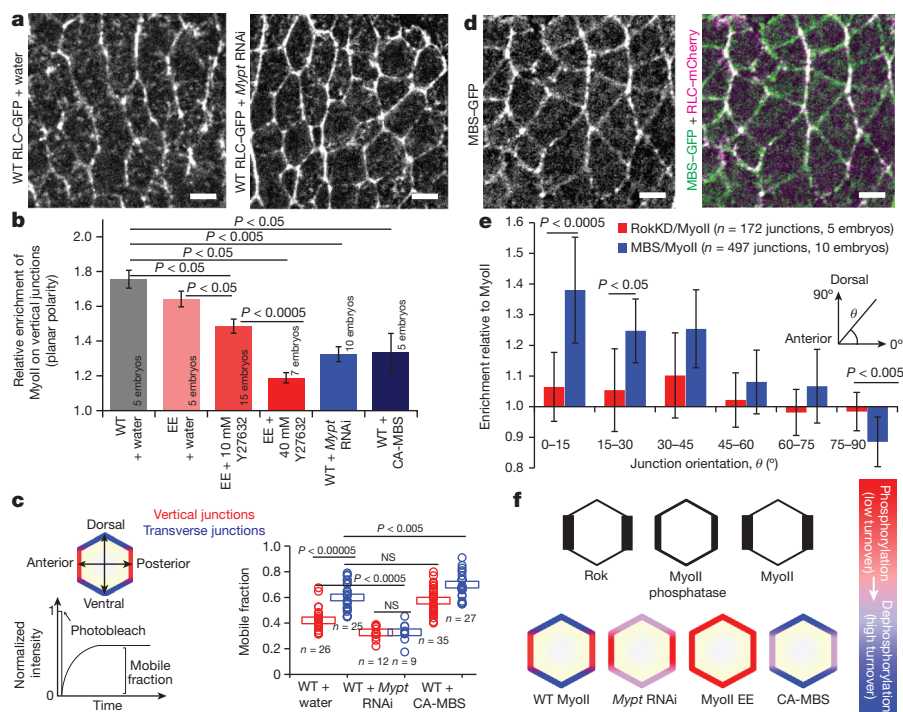


Figure 1 | Spatial control over RLC phosphorylation cycles is required for MyoII polarity. **a**, MyoII distribution in water and *Mypt-75d* dsRNA-injected embryos. **b**, Average planar polarity of MyoII in different conditions. P values calculated with Mann–Whitney U -test. **c**, Cell shows ‘vertical’ and ‘transverse’ junctions. Graph illustrates FRAP. Plot shows the mobile fraction of RLC in different conditions. Values from each junction (open circles) and the average from all junctions (open rectangle) are shown. P values calculated with Mann–Whitney U -test. NS, not significant. **d**, Distribution of MBS-GFP and RLC-mCherry. **e**, Average enrichment of kinase-dead Rok (RokKD)-GFP or MBS-GFP relative to MyoII. P values calculated with Student’s t -test. **f**, Cartoon depicts Rok, MyoII phosphatase and MyoII distribution on cell junctions and contributions of RLC phosphorylation/dephosphorylation in different experimental conditions. Scale bars, 5 μ m.

¹Aix Marseille Université, CNRS, IBDM UMR7288, 13009 Marseille, France. ²Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, Illinois 60637, USA.

and Extended Data Fig. 1d, e). A similar distribution was observed for GFP-tagged phosphomimetic RLC. Phosphomimetic RLC-GFP is not viable in the presence of the null allele of *Drosophila* RLC¹³, and was thus co-expressed with mCherry-tagged wild-type RLC (Extended Data Fig. 2a). In these embryos, a pharmacological Rok inhibitor (40 mM Y27632) specifically decreased the recruitment of wild-type RLC, while phosphomimetic RLC remained at junctions (Extended Data Fig. 2a, b) and exhibited a dose-dependent reduction of planar polarity (Fig. 1b). Diminished planar polarity in these conditions was due to higher recruitment of MyoII at transverse junctions compared to controls (Extended Data Fig. 2e). Conversely, over expression of a constitutively active mutant of MBS that lacks the target site regulated by Rok (MBSN300)¹¹, reduced MyoII levels at junctions (Extended Data Fig. 2c–e) indicative of Rok-dependent MBS inhibition on vertical junctions. Thus, steady-state levels of MyoII at junctions depend on RLC phospho-cycling¹⁸ suggesting that its polarized distribution requires Rok-dependent recruitment and MyoII phosphatase-dependent dissociation.

We tested this hypothesis by fluorescence recovery after photobleaching (FRAP). The mobile fraction from FRAP gives an estimate of the pool available for exchange (Fig. 1c). Junctional RLC exhibits lower turnover at vertical compared to transverse junctions¹⁹ (Extended Data Fig. 3a, c and Fig. 1c). Expression of the constitutively active MBS resulted in higher mobile fraction of RLC on all junctions (Fig. 1c). Mobile fraction was reduced in transverse junctions in *Mypt-75d* RNAi embryos and was lower than the mobile fraction of vertical junctions in controls (Extended Data Fig. 3b, c and Fig. 1c). Similarly, the mobile fraction of phosphomimetic RLC became progressively lower when the endogenous wild-type RLC was reduced by inhibiting Rok (Y27632; Extended Data Fig. 3d). Reduced dephosphorylation resulted in higher tension at the transverse junctions as measured by focal laser ablation^{16,19,20} (Extended Data Fig. 3e–g). Together these results show that control over the proportion and cycling of phosphorylated RLC determines the polarized stability of MyoII.

We next determined the distribution of MyoII phosphatase at cell junctions. MBS tagged with GFP²¹ (MBS-GFP; Fig. 1d) showed 30% enrichment at vertical compared to transverse junctions (Extended Data Fig. 4a). We measured the enrichment of MBS-GFP relative to MyoII and compared it to the enrichment of Rok relative to MyoII using a GFP-tagged wild-type version²² (overexpression; Extended Data Fig. 4b) and a kinase-dead version (K116A) which faithfully recapitulates Rok localization²³. Rok, like MyoII, is enriched on vertical junctions²⁴, and the ratio of wild-type Rok:RLC (Extended Data Fig. 4c) and kinase-dead Rok:RLC (Fig. 1e) was 1 at all junctions, showing positive correlation. The ratio of MBS:RLC was greater than 1 on transverse junctions and less than 1 on vertical junctions (Fig. 1e), indicative of an inverse correlation. Thus, the respective distributions of Rok and MBS regulate the stability and dissociation of MyoII at the junctions to establish planar polarity (Fig. 1f).

We next asked whether this biochemical network underlying polarity of MyoII also explains the temporal dynamics manifested in medial pulsatility. It was reported that bypassing upstream regulation of RLC abolishes MyoII pulsatility¹⁵. However, how pulses emerge from the dynamics of MyoII and its upstream regulators is unclear. Actomyosin pulses consist of two events: assembly, characterized by local increase of apical F-actin (visualized by the actin binding domain of utrophin, UtrABD-GFP)⁶ and MyoII mini-filaments concomitant with coalescence, followed by disassembly (Supplementary Video 1 and Fig. 2a). Importantly, MyoII pulses are dominated by recruitment rather than densification of mini-filaments (85%) (Fig. 2b, h), suggesting that the upstream regulators could likewise be pulsatile and entrain MyoII pulsatility. Notably, both Rok (wild-type and kinase-dead) and MBS were pulsatile with duration comparable to MyoII pulses and amplitude dominated by incorporation from the cytoplasm (75%) (Supplementary Video 2, Fig. 2c–i and Extended Data Fig. 4d).

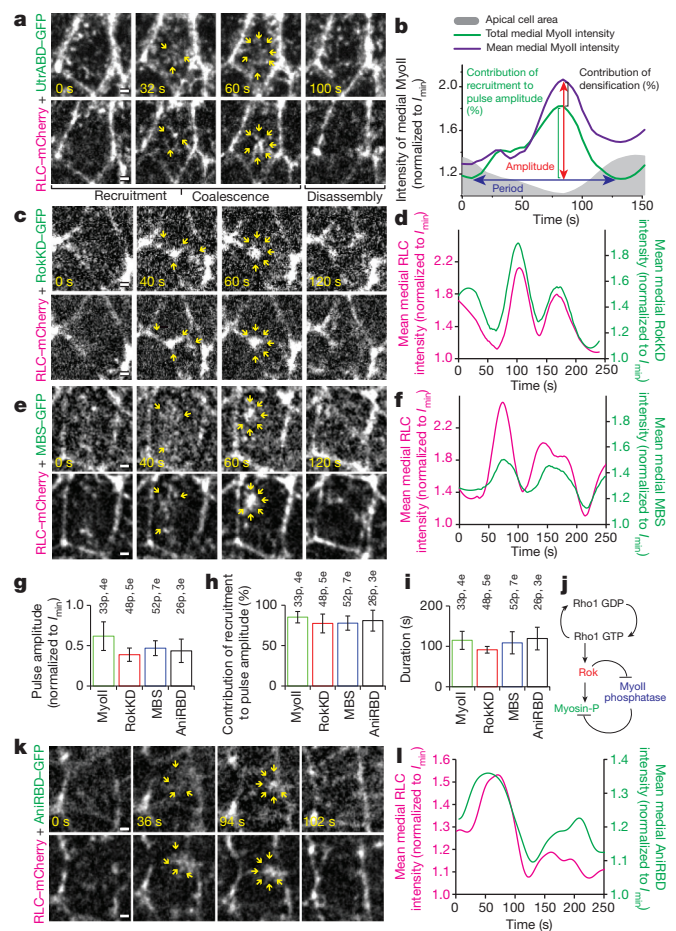


Figure 2 | MyoII pulses with its regulators Rok, MBS and Rho1 GTP. **a**, Medial F-actin densification, MyoII recruitment, coalescence and disassembly. **b**, An example of medial MyoII intensity and apical area time trace. **c**, **e**, Medial pool of kinase-dead Rok (RokKD)-GFP (**c**) and MBS-GFP (**e**) co-expressed with RLC::mCherry respectively. **d**, **f**, Corresponding medial intensity time traces. **g**, Average pulse amplitude. **h**, Average per cent contribution of recruitment to the pulse amplitude. **i**, Average duration of pulses. **j**, Biochemical pathway responsible for RLC activation (-P represents phosphorylation). **k**, Medial pool of AnirBD-GFP co-expressed with RLC-mCherry. **l**, Corresponding medial intensity time traces. **p**, number of pulses; **e**, number of embryos. Scale bars, 1 μ m.

We postulated an upstream pacemaker hypothesis whereby the pulsatility of Rok and MBS is entrained by the molecular switching of their regulator Rho1 GTPase between active-GTP and inactive-GDP forms (Fig. 2j). Indeed, Rho1 is required for MyoII activation. Injection of the C3 exoenzyme, a Rho1 GTP inhibitor²⁵, significantly reduced MyoII apical recruitment (Extended Data Fig. 5a–c). Likewise, a constitutively active Rho1 (RhoV14) led to precocious accumulation of MyoII with increased pulse amplitude and duration, and lower frequency (Supplementary Video 3 and Extended Data Fig. 5d–h). To view the activity of Rho1, we constructed a sensor by tagging the Rho1 GTP-binding domain of anillin²⁶ (AnirBD) to GFP. AnirBD-GFP was present at apical junctions, exhibited 20% enrichment on vertical junctions (data not shown) and was pulsatile (Supplementary Video 4 and Fig. 2g–l). C3 exoenzyme injection led to a significant reduction of AnirBD-GFP at junctions and loss of medio-apical pulsatility (Extended Data Fig. 5a–c) confirming it to be a sensor of Rho1 GTP.

We tested the upstream pacemaker model by doing cross-correlation analysis between medial intensity time traces of RLC-mCherry and either kinase-dead Rok-GFP, MBS-GFP or AnirBD-GFP (Fig. 3a). The correlation was strong in all cases (Pearson's correlation coefficients were 0.8 for Rok, 0.7 for MBS and 0.5 for AnirBD). While

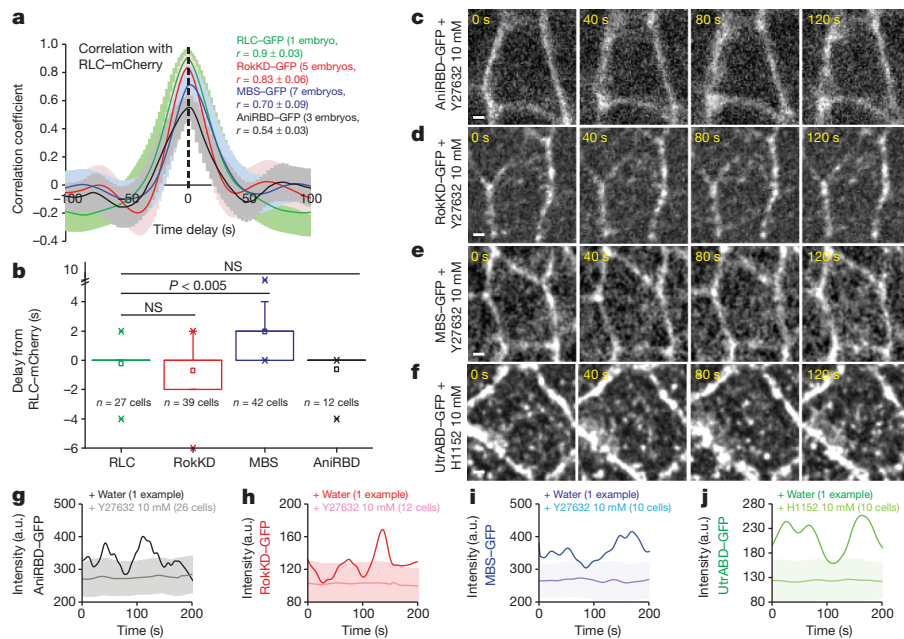


Figure 3 | Pulsatility of Rho1 GTP, Rok and MBS requires MyoII activity. **a**, Average temporal cross-correlation of RLC-mCherry with GFP-tagged upstream regulators. Pearson's correlation coefficients (r) are shown. **b**, Box plot of the time delays calculated from the cross-correlation analysis. NS, not significant. **c–f**, Time snapshots of different embryos injected with Rok inhibitors (Y27632/H1152). **g–j**, Mean medial intensity time traces of various embryos injected with Rok inhibitor and control. Error bars, s.d.; P values calculated with Mann-Whitney U -test. Scale bars, 1 μ m.

the intensity peak of MyoII concurred with Rok and AniRBD, MBS peaks followed MyoII peaks with an average delay of 2 s (Fig. 3b). The temporal delay of MBS raised the possibility that the biochemical network is entrained by Rho1 to establish pulsatility. Importantly, this predicts that Rho1 GTP pulses are independent of the activities of downstream effectors. However, inhibiting MyoII by inhibition of Rok (10 mM Y27632 or 10 mM H1152) completely blocked the cortical recruitment of AniRBD-GFP, kinase-dead Rok-GFP, wild-type Rok-GFP and MBS-GFP, and the pulsatility of cortical F-actin (Supplementary Video 5, Fig. 3c–j and Extended Data Fig. 4d). These observations are inconsistent with an upstream pace-making activity of Rho1 and suggest that pulsatility is instead an emergent property of the whole network.

Exploring the initiation of pulses, we found that early GBE (stage 5a) is associated with fluctuations of MyoII, Rok and AniRBD (Extended Data Figs 5e and 6a, b). Over time (stage 7), the fluctuations assemble to form pulses with a nonlinear rate of change (Extended Data Fig. 6c). Lack of a detectable time delay between MyoII and AniRBD fluctuations (Extended Data Fig. 6d) and non-predictable pulse position (Extended Data Fig. 6e) suggests that pulsatility involves amplification of Rho1-mediated stochastic activation. To understand the nature of this amplification, we focused on the cortical dynamics of the network components. During pulse formation, contractile MyoII speckles move in the plane of the cortex together with F-actin⁶, a process hereafter referred to as advection. Advection and concentration of MyoII, potentially with molecules that activate MyoII, could create a positive feedback that amplifies recruitment of the network components (Fig. 4p). We first measured advection using particle image velocimetry (PIV)²⁷ (Supplementary Video 6 and Fig. 4a). The velocity vectors from AniRBD, Rok, MBS and F-actin strongly correlated with vectors of MyoII (Fig. 4a, b), showing that these are advected together. We next calculated the divergence of the vector fields. Actomyosin networks, Rho1, Rok and MBS converge (negative divergence) in the centre of the cell during pulse assembly (Fig. 4c, d). MyoII contractility drives advection as inhibition of MyoII activity blocks F-actin convergence (Fig. 4g, h).

To test the requirement of convergent advection for pulse assembly we injected a low dose of cytochalasin D (1 mg ml⁻¹) to disrupt actin polymerization. Although the effect on F-actin density was mild, MyoII convergence was reduced (Supplementary Video 7 and Fig. 4e, g, h). Consistent with convergent advection driving signal amplification

(by locally concentrating upstream regulators), the amplitudes of both Rok and MyoII pulses were reduced (Fig. 4f). We next sought to disentangle the contribution of advection and biochemical recruitment. Pairwise comparison of RLC pulses with Rho1/kinase-dead Rok pulses showed that their amplitudes had a linear relation, suggesting feedback between advection and biochemical recruitment (Extended Data Fig. 7c). A Rok mutant lacking the Rho1-binding domain (kinase-dead Rok Δ ARB)²⁸ was pulsatile through advection, but exhibited a lower pulse amplitude (Extended Data Fig. 7a, b) and lost the linear relation with MyoII pulses (Extended Data Fig. 7c). Thus, convergent advection promotes concentration and recruitment by biochemical activation resulting in positive feedback. Cytosolic proteins (GFP) or low-affinity actin-binding proteins (α -catenin) are not pulsatile (Extended Data Fig. 7d, e), indicating that pulsatility requires specific recruitment to actomyosin networks.

We next addressed pulse disassembly and the potential role of RLC dephosphorylation. *Mypt-75d* RNAi and intermediate inhibition of wild-type RLC in phosphomimetic -RLC-expressing embryos (10 mM Y27632) (Supplementary Video 8 and Fig. 4i) increased pulse amplitude and duration, and reduced frequency (Extended Data Fig. 8a–c), similar to RhoV14 expression (Extended Data Fig. 5f–h). Conversely, expression of constitutively active MBS, led to decreased amplitude and duration of MyoII pulses (Fig. 4k, Extended Data Fig. 8a–c). Interestingly, reduced dephosphorylation also lowered the advection velocities of MyoII and Rok (Extended Data Fig. 8d). A complete block of phospho-cycling in the phosphomimetic RLC mutant (40 mM Y27632) led to failure of convergent advection, thus entirely abolishing pulsatility (Supplementary Video 8, Fig. 4j, l and Extended Data Fig. 8e, f). We conclude that RLC dephosphorylation mediates dissociation of MyoII during pulse disassembly and sets the regime whereby MyoII can be advected with its regulators and thereby be pulsatile.

However, a constant ratio of MyoII recruitment and advection, and dissociation cannot account for MyoII concentration changes in a pulse (Extended Data Fig. 9a). Pulse disassembly requires a negative feedback signal that could stem either from a gradual decline of advection up to a point where dissociation supersedes advection, and/or a delayed increase of MyoII dissociation rate²⁹ (as observed in Fig. 3b). We estimated MyoII turnover during pulsatility by FRAP in embryos co-expressing wild-type RLC-mCherry and wild-type RLC-GFP (Extended Data Fig. 9b, e). We photo-bleached the mCherry channel and normalized the recovery time trace to the GFP channel (Extended

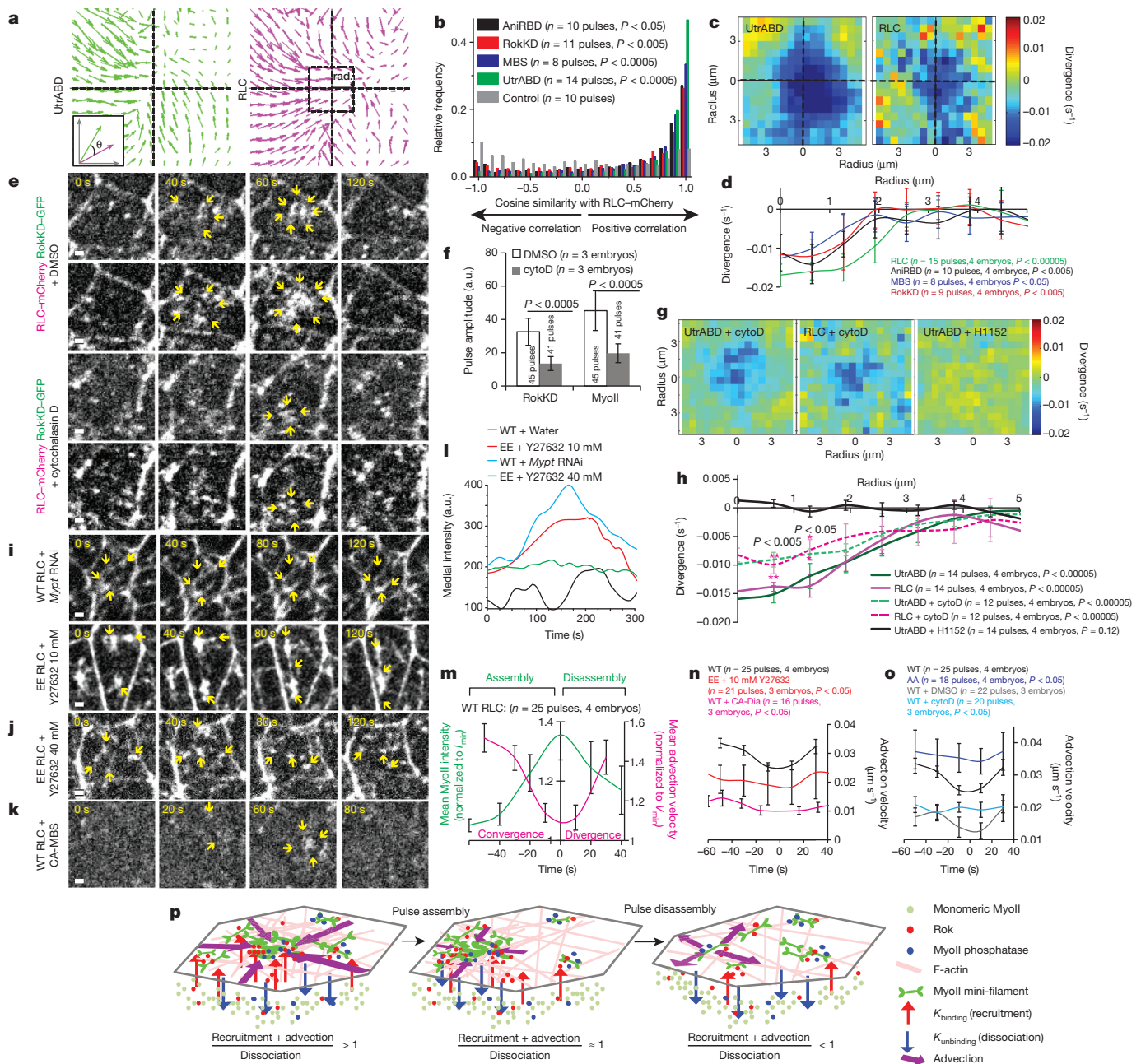


Figure 4 | Pulsatility is an emergent behaviour of a biomechanical network.

a, Vector fields of UtrABD and RLC averaged over pulse assembly. Inset illustrates the principle of cosine similarity. Dotted box illustrates the perimeter of a square for calculating divergence as a function of radius (rad.). **b**, Distribution of cosine similarity indices from average vector fields. P values calculated with Kolmogorov–Smirnov test. **c**, Heat map of average divergence values as a function of radius. P values calculated with one-sample t -test. **d**, Plot shows average divergence values as a function of radius. P values calculated with one-sample t -test. **e**, Time snapshots of embryos co-expressing RLC–mCherry and kinase-dead Rok (RokKD)–GFP injected either with DMSO or cytochalasin D. **f**, Average pulse amplitudes in these conditions. P values calculated with Mann–Whitney U -test. **g**, Average divergence values in embryos injected with either cytochalasin D or

H1152. **h**, Plot shows average divergence values as a function of radius in various conditions. P values in brackets: one-sample t -test; P values above asterisks: Mann–Whitney U -test. **i**, Time snapshots of wild-type embryos injected with *Myot* dsRNA and phosphomimetic (EE) embryos injected with 10 mM Y27632. **j**, Time snapshots of phosphomimetic embryos injected with 40 mM Y27632. **k**, Time snapshots of wild-type embryos expressing constitutively active MBS. **l**, Representative medial intensity time traces in these conditions. **m**, Mean medial MyoII intensity and the corresponding advection velocities. **n**, **o**, Average advection velocities of MyoII in different conditions as a function of time. P values calculated with Mann–Whitney U -test. **p**, Cartoon model of pulse assembly and disassembly. Scale bars, 1 μ m.

Data Fig. 9c, f). During assembly the ratio of mCherry and GFP channels remained constant (though at different values) pre- and post-bleaching, indicating low turnover (Extended Data Fig. 9b–d, h). During disassembly, however, the ratio recovered to $\sim 50\%$ post-bleaching, suggesting a higher turnover (Extended Data Fig. 9e–h). In addition, we found that MyoII advection rates gradually declined and reached a minimum when MyoII intensity reached a maximum (Fig. 4m). What causes this reduction of advection? As in junctions

(Extended Data Fig. 3f, g), recruitment and densification of medial actomyosin networks could increase local friction or tension, thereby reducing advection. Indeed, the advection velocity was reduced when medial MyoII concentration was high (Extended Data Fig. 8b and Fig. 4n), or when F-actin networks were made denser by the expression of a constitutively active form of a Rho1-activated actin nucleator termed ‘diaphanous’ (lacking the domains responsible for intramolecular inhibition)’ (Supplementary Video 9 and Fig. 4n). Conversely,

expression of phospho-dead RLC (T20A/S21A) or cytochalasin D injection (1 mg ml^{-1}) caused no reduction of advection compared to controls (Fig. 4o). We conclude that MyoII advection is reduced due to densification of actomyosin networks. Notably, MyoII advection began to recover during disassembly, though it was associated with divergence (Supplementary Video 6, Fig. 4m and Extended Data Fig. 9i, j), suggesting mechanical relaxation of the actomyosin networks amplifying Rho1, Rok and MyoII dispersion and depletion.

Under the regulation of Rho1 activity, actomyosin networks exhibit three classes of behaviours (Extended Data Fig. 10). When Rho1 activity is high, MyoII stably binds F-actin and increases network tension (vertical junctions or when RhoV14/phosphomimetic RLC is expressed or phosphatase is inhibited). When Rho1 activity is low, MyoII dissociation is high, resulting in lower network tension (transverse junctions or when constitutively active MBS/phospho-dead RLC is expressed). Between these two behaviours, a pulsatile regime emerges as a result of feedback between advection and dissociation rates: advection-mediated positive feedback amplifies Rho1, Rok and MyoII recruitment in a cascade; but actomyosin concentration negatively feeds back to reduce advection. Such self-organization of biomechanical networks could operate in other systems where actomyosin flows have been observed, such as apical cell constriction, cell migration and cytokinesis⁹. Spatial control over MyoII turnover thus provides a potentially general mechanism to drive polarized shape changes during tissue morphogenesis. How polarity of these mechanical ratchets is achieved in different tissues still remains unclear. We propose that the intrinsic state of actomyosin networks and their regulators is a self-organized biomechanical process, which is tuned by Rho1 activity, regulated in a tissue-dependent manner and fundamental to morphogenesis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 September 2014; accepted 29 May 2015.

Published online 27 July 2015.

- Munro, E., Nance, J. & Priess, J. R. Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior-posterior polarity in the early *C. elegans* embryo. *Dev. Cell* **7**, 413–424 (2004).
- Kim, H. Y. & Davidson, L. A. Punctuated actin contractions during convergent extension and their permissive regulation by the non-canonical Wnt-signaling pathway. *J. Cell Sci.* **124**, 635–646 (2011).
- Martin, A. C., Kaschube, M. & Wieschaus, E. F. Pulsed contractions of an actin-myosin network drive apical constriction. *Nature* **457**, 495–499 (2009).
- Solon, J., Kaya-Copur, A., Colombelli, J. & Brunner, D. Pulsed forces timed by a ratchet-like mechanism drive directed tissue movement during dorsal closure. *Cell* **137**, 1331–1342 (2009).
- He, L., Wang, X., Tang, H. L. & Montell, D. J. Tissue elongation requires oscillating contractions of a basal actomyosin network. *Nature Cell Biol.* **12**, 1133–1142 (2010).
- Rauzi, M., Lenne, P. F. & Lecuit, T. Planar polarized actomyosin contractile flows control epithelial junction remodelling. *Nature* **468**, 1110–1114 (2010).
- Kaibuchi, K., Kuroda, S. & Amano, M. Regulation of the cytoskeleton and cell adhesion by the Rho family GTPases in mammalian cells. *Annu. Rev. Biochem.* **68**, 459–486 (1999).
- Bertet, C., Sulak, L. & Lecuit, T. Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* **429**, 667–671 (2004).
- Levayer, R. & Lecuit, T. Biomechanical regulation of contractility: spatial control and dynamics. *Trends Cell Biol.* **22**, 61–81 (2012).
- Niederman, R. & Pollard, T. D. Human platelet myosin. II. *In vitro* assembly and structure of myosin filaments. *J. Cell Biol.* **67**, 72–92 (1975).
- Lee, A. & Treisman, J. E. Excessive Myosin activity in mbs mutants causes photoreceptor movement out of the *Drosophila* eye disc epithelium. *Mol. Biol. Cell* **15**, 3285–3295 (2004).
- Kimura, K. *et al.* Regulation of myosin phosphatase by Rho and Rho-associated kinase (Rho-kinase). *Science* **273**, 245–248 (1996).
- Jordan, P. & Karess, R. Myosin light chain-activating phosphorylation sites are required for oogenesis in *Drosophila*. *J. Cell Biol.* **139**, 1805–1819 (1997).
- Dawes-Hoang, R. E. *et al.* folded gastrulation, cell shape change and the control of myosin localization. *Development* **132**, 4165–4178 (2005).
- Vasquez, C. G., Tworoger, M. & Martin, A. C. Dynamic myosin phosphorylation regulates contractile pulses and tissue integrity during epithelial morphogenesis. *J. Cell Biol.* **206**, 435–450 (2014).
- Kasza, K. E., Farrell, D. L. & Zallen, J. A. Spatiotemporal control of epithelial remodeling by regulated myosin phosphorylation. *Proc. Natl Acad. Sci. USA* **111**, 11732–11737 (2014).
- Vereshchagina, N. *et al.* The essential role of PP1beta in *Drosophila* is to regulate nonmuscle myosin. *Mol. Biol. Cell* **15**, 4395–4405 (2004).
- Watanabe, T., Hosoya, H. & Yonemura, S. Regulation of myosin II dynamics by phosphorylation and dephosphorylation of its light chain in epithelial cells. *Mol. Biol. Cell* **18**, 605–616 (2007).
- Fernandez-Gonzalez, R., Simoes Sde, M., Roper, J. C., Eaton, S. & Zallen, J. A. Myosin II dynamics are regulated by tension in intercalating cells. *Dev. Cell* **17**, 736–743 (2009).
- Rauzi, M., Verant, P., Lecuit, T. & Lenne, P. F. Nature and anisotropy of cortical forces orienting *Drosophila* tissue morphogenesis. *Nature Cell Biol.* **10**, 1401–1410 (2008).
- Sen, A., Nagy-Zsver-Vadas, Z. & Krahn, M. P. *Drosophila* PATJ supports adherens junction stability by modulating Myosin light chain activity. *J. Cell Biol.* **199**, 685–698 (2012).
- Bardet, P. L. *et al.* PTEN controls junction lengthening and stability during cell rearrangement in epithelial tissue. *Dev. Cell* **25**, 534–546 (2013).
- Mason, F. M., Tworoger, M. & Martin, A. C. Apical domain polarization localizes actin-myosin activity to drive ratchet-like apical constriction. *Nature Cell Biol.* **15**, 926–936 (2013).
- Simões Sde, M. *et al.* Rho-kinase directs Bazooka/Par-3 planar polarity during *Drosophila* axis elongation. *Dev. Cell* **19**, 377–388 (2010).
- Nemoto, Y., Namba, T., Kozaki, S. & Narumiya, S. Clostridium botulinum C3 ADP-ribosyltransferase gene. Cloning, sequencing, and expression of a functional protein in *Escherichia coli*. *J. Biol. Chem.* **266**, 19312–19319 (1991).
- Piekny, A. J. & Glotzer, M. Anillin is a scaffold protein that links RhoA, actin, and myosin during cytokinesis. *Curr. Biol.* **18**, 30–36 (2008).
- Levayer, R. & Lecuit, T. Oscillation and Polarity of E-Cadherin Asymmetries Control Actomyosin Flow Patterns during Morphogenesis. *Dev. Cell* **26**, 162–175 (2013).
- Simões Sde, M., Mainieri, A. & Zallen, J. A. Rho GTPase and Shroom direct planar polarized actomyosin contractility during convergent extension. *J. Cell Biol.* **204**, 575–589 (2014).
- Lewis, J. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.* **13**, 1398–1408 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Rao for useful discussions, and the Lecuit and Lenne groups and the Labex INFORM (ANR-11-LABX-0054) for support and discussions. We are grateful to M. Krahn, A. Martin, V. Mirouse, J. Treisman and J. Zallen for the gift of flies. This work was supported by the ANR Archiplast and the ERC (Biomecamorph #323027) and HFSP (T.L. and E.M.). A.M. was supported by Ministère de l'Éducation nationale and the Association pour la Recherche contre le Cancer (ARC). We acknowledge the France-Biomedicine/PICSL infrastructure (ANR-10-INSB-04-01).

Author Contributions This project originated from discussions between T.L. and E.M., based on a computational model developed by E.M. A.M. and T.L. planned the project, discussed the data with input from E.M., and wrote the paper. A.M. did all the experiments and analysis. J.-M.P. made the fly constructs and helped with molecular biology.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L. (thomas.lecuit@univ-amu.fr).

METHODS

Transgenic Lines. MyoII RLC is encoded by the gene *spaghetti-squash* (*sqh*, Genebank ID: AY122159)³⁰. For generating GFP-tagged phosphomimetic Sqh (phosphomimetic RLC–GFP) and phospho-dead Sqh (phospho-dead RLC–GFP), *sqh-mCherry* plasmid under its native promoter (kind gift from A. Martin)³ was modified by replacing mCherry (ATG→TAA) to monomeric eGFP (A206→K). Corresponding codons of the amino acids Thr20 and Ser21 on the *sqh* gene were modified to Glu20 and Glu21 (phosphomimetic) or Ala20 and Ala21 (phospho-dead) using the Quick-Change Mutagenesis Kit (Stratagene). Recombinant expression vectors were sent to BestGene Inc. for random insertion transgenesis.

To construct the Rho1 GTP sensor (AniRBD–GFP), monomeric eGFP was fused to the C-terminal end of *Drosophila* anillin (amino acids 748–1239, Genebank ID: AAL39665) that lacks the N-terminal myosin- and actin-binding domain but retains its Rho1-binding domain. A GSSGSS linker was added in between GFP and anillin. This sensor is similar to the one used in human cells³⁶. The transgene is expressed under the ubiquitous p63E promoter. For cloning, *Ubi-GFP::AniRBD* ORF replaced the *stop-CD8GFP* in the *Ubi-stop-mCD8GFP* plasmid obtained from Stepan Belyakin, Russia (GeneBank ID: KC845568). Recombinant expression vector was sent to BestGene Inc. for PhiC31 site-specific mediated insertion transgenesis.

All FASTA sequences of the complete vectors are available on request.

Fly constructs and genetics. Fly (*Drosophila melanogaster*) constructs and genetics are listed below, where asterisks denote UAS (upstream activating sequence) males crossed with maternal transcription *67gal4*-expressing females (matttbgal4VP16).

Wild-type RLC–GFP distribution and dynamics in various conditions. (1) *yw cv sqhAX3;sqh-sqh::GFP*;+ (a gift from R. Karess) for water, *Mypt* dsRNA, cytochalasin D and DMSO injections and PIV in Figs 1a–c, 2b, 4i, l, m–o, Extended Data Figs 1c, 2e, 3, 6e, 8a–c, e, f, 9i, j and Supplementary Videos 6 and 8. (2) F2 progeny*: male *UAS-MBS shRNA* crossed with female +;67Gal4,*sqh-sqh::GFP*;+ (*UAS-MBSshRNA* is a Bloomington Stock 41625) in Extended Data Fig. 1d, e. (3) F2 progeny*: male *UAS-CA-MBS* crossed with female +;67Gal4,*sqh-sqh::GFP*;+ (*UAS-CA-MBS/MBSN300* was a gift from J. Treisman) in Fig. 4k and Extended Data Figs 2c–e and 8a–c. (4) F1 progeny*: male *UAS-RhoV14* crossed with female +;67Gal4,*sqh-sqh::GFP*;+ (*UAS-RhoV14* is a Bloomington Stock 7330) in Extended Data Fig. 5d, e and Supplementary Video 3. (5) F1 progeny*: male *UAS-diaCA NC* crossed with female +;67Gal4,*sqh-sqh::GFP*;+ (*UAS-diaCA NC* was a gift from P. Rorth) in Figure 4n and Supplementary Video 9.

Phosphomimetic RLC (EE–RLC)–GFP-expressing embryos. Phosphomimetic-RLC-expressing embryos: *sqh-sqhEE::GFP*;+;+ (see transgenic lines) compared to wild-type embryos: +;67Gal4,*sqh-sqh::GFP*;+ (a gift from R. Karess) in Fig. 1b, 4i, j, l, n, Extended Data Figs 2e, 3d–g, 8a–c, e, f and Supplementary Video 8.

Rok inhibition in embryos co-expressing phosphomimetic RLC–GFP and wild-type RLC–mCherry. +;67Gal4,*sqh-sqh::GFP*,*sqh-sqh::Cherry*;+ compared with *sqh-sqhEE::GFP*; +;67Gal4,*sqh-sqh::Cherry*;+ (*sqh-sqh::Cherry* was a gift from A. Martin) in Extended Data Fig. 2a, b.

FRAP of phosphomimetic RLC–GFP- and wild-type RLC–GFP-expressing embryos. Four copies: +;67Gal4,*sqh-sqh::GFP*;+ compared to *sqh-sqhEE::GFP*;+;+ in Extended Data Fig. 2e. Two copies: *yw cv sqhAX3;sqh-sqh::GFP*;+ compared to *sqh-sqhEE::GFP*;+;+ injected with 40 mM Y27632 in Extended Data Fig. 2e.

UtrABD::GFP dynamics. *yw cv sqhAX3;sqh-UtrABD::GFP/Cyo;sqh-sqh::mCherry* in Figs 2a, 3f, j and 4a–c, g, h. F1 progeny*: male *UAS-diaCA NC* crossed with female +;67Gal4/*sqh-UtrABD::GFP;sqh-sqh::mCherry*/+ in Supplementary Video 9.

MBS–GFP distribution and dynamics. F2 progeny*: male *UAS-MBS::GFP* crossed with female *yw cv sqhAX3;67Gal4, sqh-sqh::Cherry*;+ (*UAS-MBS::GFP* was a gift from M. Krahn) in Figs 1d, e, 2e–i, 3a, b, e, i, 4b, d, Extended Data Fig. 4a and Supplementary Videos 2 and 5.

RokKD–GFP distribution and dynamics. (1) +;67Gal4,*sqh-sqh::mCherry;sqh-RokKD::GFP* (*sqh-RokKD::GFP* was a gift from J. Zallen) in Figs 1e, 2c, d, g–i, 3a, b, d, h, 4b, d–f, Extended Data Figs 6a, c, 7c, 8d and Supplementary Videos 2 and 5. (2) F1 progeny*: male *UAS-RhoV14*; crossed with female *67Gal4, sqh-RokKD::GFP*/+; *sqh-sqh::Cherry*/+ (*sqh-RokKD::GFP* was a gift from J. Zallen) in Extended Data Fig. 5f–h.

Wild-type Rok–GFP distribution and dynamics. +;67Gal4,*sqh-sqh::Cherry;Ubi-Rok::GFP/Tm6* (*Ubi-Rok::GFP* was a gift from V. Mirouse) in Extended Data Fig. 4b–d.

Rho1 GTP sensor distribution and dynamics. +;67Gal4,*sqh-sqh::Cherry*;+ (see transgenic lines) in Figs 2g–l, 3a–c, g, 4b, d, Extended Data Figs 5a–c, 6b–d, 7c and Supplementary Videos 4 and 5.

Cytosolic GFP dynamics. F1 progeny*: male *UAS-GFP* crossed with female *yw cv sqhAX3;67Gal4, sqh-sqh::Cherry*;+ (*UAS-GFP* is a Bloomington Stock 30140) in Extended Data Fig. 7d, e.

α -Catenin–YFP dynamics. +;67Gal4,*sqh-sqh::mCherry* (α -Catenin::YFP is a Kyoto Stock) in Extended Data Fig. 7d, e.

RokKD Δ ABD–Venus and control dynamics. F2 progeny*: male *UAS-RokK116A* crossed with female *yw cv sqhAX3;67Gal4, sqh-sqh::Cherry*;+; and male *UAS-RokK116A Δ ABD* crossed with *yw cv sqhAX3;67Gal4, sqh-sqh::Cherry*;+ (gifts from J. Zallen) in Extended Data Fig. 7a–c.

Phospho-dead RLC (AA–RLC)–GFP-expressing embryos. +;67Gal4,*sqh-sqhAA::GFP* compared with: +;67Gal4,*sqh-sqh::GFP*;+ (see transgenic lines) in Fig. 4o.

Two-colour FRAP. +;67Gal4,*sqh-sqh::GFP*,*sqh-sqh::Cherry*;+ in Extended Data Fig. 9b–h.

RNA interference. A double-stranded RNA probe against *Mypt-75d* was made using PCR product containing the sequence of the T7 promoter (TAATACG ACTCACTATAGGG) followed by 21 nucleotides specific to the gene. Sequences were amplified between nucleotides 3,133–3,517 (GeneBank ID: AY075426). PCR products were subsequently used as a template for the *in vitro* RNA synthesis. 0–1-h-old embryos were injected using 5 μ M dsRNA diluted in RNase-free water.

For knockdown of MBS, UAS-driven small hairpin RNA (shRNA) was expressed as described in fly constructs and genetics.

Real-Time PCR. RNAi efficiency was estimated by measuring endogenous mRNA levels using quantitative RT–PCR after injection with dsRNA probe against *Mypt-75d*. Total RNA extraction from early gastrulating embryos was performed using the miRNeasy micro kit (QIAGEN, no. 217084) with an additional DNase treatment step to remove genomic DNA contamination. Reverse transcription was performed with the iScript Reverse Transcription Supermix (Bio-RAD, no. 170-8841) according to the manufacturer's protocol. Real-time PCR was performed on a CFX96 QPCR detection system (Bio-RAD) with the following TaqMan probes (TaqMan Gene Expression Assays from Life Technologies) plus TaqMan Gene expression master mix, following classical TaqMan protocol:

- MYPT-75D: Dm01806472_g1 = probe 1 (exon 7–8 boundary, 1983/GenBank AY075426, amplicon = 142 bp) and Dm01806468_g1 = probe 2 (exon 3–4 boundary, 948/GenBank AY075426, amplicon = 98 bp).
- House-keeping gene referent RPL32: Dm02151827_g1 (exon 2–3 boundary, 948/GenBank AY070656, amplicon = 72 bp)
- House-keeping gene referent ACT42A: Dm02362162_s1 (exon 2, 948/GenBank AY118907, amplicon = 108 bp)
- House-keeping gene referent RPII140: Dm02134593_g1 (exon 2–3 boundary, 2331/GenBank BT003265, amplicon = 78 bp)

Quantitative RT–PCR conditions were as follows: 40 cycles of 95 °C for 15 s and 60 °C for 60 s. Analyses were performed in triplicate from three independent experiments. After comparison of three house-keeping genes, transcript levels were first normalized to one, *Rpl-32*; and then to their respective control group. $\Delta\Delta C_q$ method was used to estimate relative amounts using the Bio-RAD CFX manager software.

A threefold depletion of *Mypt-75d* transcripts was observed in *Mypt-75d* RNAi embryos compared to control embryos with both Taqman probes. *P* values were calculated with unpaired *t*-tests using qBasePLUS software (Biogazelle).

Drug injections. Y27632 dihydrochloride (Sigma) was injected in stage 5a embryos expressing phosphomimetic RLC at 40 mM for strong inhibition, and at 10 mM for weaker inhibition of endogenous RLC or wild-type RLC–mCherry. For wild-type embryos, inhibition was strong at 10 mM concentration. Inhibition of MyoII by Y27632 (10 mM) was comparable to another Rok inhibitor, namely H-1152 (Enzo Life Science) at 10 mM.

C3 exoenzyme (Enzo Life Sciences) was injected into stage 5a embryos at 2 μ g μ l^{−1}.

Cytochalasin D (Sigma) suspended in DMSO was injected into stage 5a embryos at 1 mg ml^{−1} concentration diluted from a 10 mg ml^{−1} stock. Corresponding dilution of DMSO was made for control injections.

Image acquisition. Embryos were prepared as described before³¹. Time-lapse imaging was done from stage stage 5a for observing medial fluctuations during early GBE and from stage 7 to stage 8 for all other experiments (5–30 min depending on the experiment). A Nikon spinning disc Eclipse Ti inverted microscope using a 100 \times , 1.4 N.A oil-immersion objective at room temperature (~22 °C) was used. The system acquires images using the MetaMorph software. For junctional and medial intensity measurements, starting from the most apical plane, seven *z*-sections, 0.5 μ m apart, were acquired every 4 or 5 s using a single camera. For performing temporal cross-correlation analysis on medial intensity time traces, three apical *z*-sections 0.75 μ m apart were acquired every 2 s using a single camera. For visualizing UtrABD–GFP, only three apical *z*-sections 0.5 μ m apart were acquired. Laser power was measured and kept constant between experiments.

FRAP was performed using the targeted laser of the Nikon spinning disc (iLAS²). A constant junctional area was photo-bleached (0.91 μ m diameter) for every experiment.

Laser ablation. Nano-ablation of junctions was performed on an inverted microscope (Eclipse TE 2000-E, Nikon) with a spinning disk (Ultraview ERS, Perkin Elmer) using a 100 \times , 1.4 N.A. oil-immersion objective. A home-built system was used as described previously²⁰, using a focused laser beam exposed for 5–10 ms with a power between 800–950 mW depending on the experiment. Images were acquired every 200 ms to get a better temporal resolution.

Image analysis. All image processing was done in Fiji freeware software. For all quantifications, sum-intensity z -projection of slices was used, followed by a background subtraction using the available plugin in Fiji ('subtract background', ball radius 30 pixels, 1 pixel = 0.13 μ m).

For junctional intensity and planar polarity measurements, regions were drawn manually (width 6 pixels). Junctions were categorized based on angles (0–15° corresponding to transverse and 75–90° corresponding to vertical as shown in Fig. 1c). For plotting planar polarity, intensity from each vertical junction was normalized to the mean of transverse junctions in a given embryo. These values were averaged per embryo. The means of the averaged values were taken to plot Fig. 1b, Extended Data Fig. 1e and Extended Data Fig. 4a. For comparing junctional intensity values in Extended Fig. 2, intensity from each junction was normalized to the mean of the junctions of the respective control embryos. The means of the normalized values were plotted in all graphs of Extended Data Fig. 2.

For pulsatility measurements, individual cells were tracked using CellTrack1.1 (open source software link: <http://bio.cse.ohio-state.edu/CellTrack/>). The software tracks cells using vertices as points of reference. The procured region was shrunk using a constant thickness (6 pixels) to exclude contribution from junctional signals. Medial intensity time traces were smoothed to automatically calculate pulse amplitude and duration using Igor Pro (Wavemetrics). To calculate normalized pulse amplitude and per cent recruitment of MyoII and upstream regulators (Fig. 2g, h), mean intensity and total intensity time traces were normalized to their minimum values respectively (base value of pulses) and treated as described above to extract amplitudes. Values plotted in Fig. 2g are amplitudes from normalized mean intensity time traces. Total intensity pulse amplitudes were subtracted from the respective mean intensity pulse amplitudes to calculate the per cent contribution of densification as a result of apical area reduction (as depicted in Fig. 2b). Pulse amplitudes reported in all other figures are calculated from mean intensity time traces.

The first derivative of increase in intensity during pulse assembly was calculated in Igor using normalized and smoothed mean intensity time traces (Extended Data Fig. 6c).

To measure the location of pulses, the Fiji function 'centre of mass' was used to get the coordinates of pulse coalescence in a region of interest with constant size (60 by 60 pixels, Extended Data Fig. 6e).

FRAP quantification. Bleached region was tracked using a home-made Fiji macro. Measurements were done only on those intensity time traces that reached a plateau. Intensity at each time point was normalized as follows to scale the intensity time trace between 0 and 1

$$\frac{I_t - I_{\min}}{I_{\max} - I_{\min}} = y$$

Where: I_{\max} , pre-bleach intensity; I_{\min} , bleached intensity; I_t , intensity at time t ; and y , normalized intensity.

The normalized recovery trace was fit to a single exponential in Origin (version 8) to obtain the mobile fraction as follows

$$y = A(1 - e^{-t/\tau})$$

Where: A , mobile fraction (end value of the recovered intensity); τ , fitted parameter; t , time after bleaching.

In two-colour FRAP experiments, the ratio of medial MyoII in red and green channels was used to do the above normalization (Extended Data Fig. 9b–h).

Laser ablation quantification. Recoil of the junction after ablation was measured by tracking its vertices for 2 s. A linear fit was used to calculate the recoil velocity.

Temporal cross-correlation analysis. Temporal cross-correlation of medial intensity time traces from RokKD, MBS and AnirBD with MyoII RLC were done

in Igor Pro, as described previously²⁷, to obtain the Pearson's correlation coefficient (r) and time delay.

PIV analysis. PIV analysis was performed using the MATPIV toolbox in MATLAB as described previously²⁷, and the vectors were averaged over time to obtain average vector fields. The analysis was restricted to a central region of the cell. The ROI was usually 60 pixels by 60 pixels (1 pixel = 0.13 μ m) to catch significant medial pool and to avoid perturbation from the junctional signal. For PIV of UtrABD–GFP + RLC–mCherry, images were rescaled to get a pixel size of 0.16 μ m (instead of 0.08 μ m). For comparing the direction of the vectors in two channels (as depicted in Fig. 4a), the cosine similarity index was calculated from vector fields averaged over one pulse cycle (100–120 s) for each channel (red (R) or green (G)) as follows

$$\text{Cosine similarity} = \frac{R \cdot G}{\|R\| \|G\|} = \frac{R_i \times G_i}{\sqrt{R_i^2} \sqrt{G_i^2}}$$

The resultant value ranges between -1 (opposite angles) to 0 (independent angles) to $+1$ (same angle). As a control, images were taken from sub-apical sections (1.5 μ m deep) of red (RLC–mCherry) and green (UtrABD–GFP) channels to estimate the contribution of junctional drift to the similarity index.

For calculating divergence, averaging of displacement vectors was done either during pulse assembly (duration 0–~60 s) or disassembly (~90–120 s). MATLAB 'div' function was used to calculate divergence from the average vector fields. The heat maps in Fig. 4c, g and Extended Data Figs 8e and 9i are average divergence values from many cells (10 μ m \times 10 μ m). The divergence in Fig. 4d, h and Extended Data Figs 8f and 9j were calculated by averaging the divergence values on a perimeter of a square expanding from the centre (0) up to a radius of 5 μ m (as depicted in Fig. 4a).

For calculating advection velocities during the course of the pulse, PIV was performed on a pulse cycle divided into bins of 20 s. Integration of displacement vectors was done for each bin (20 s), the norms of the resulting vectors were averaged in a given region of interest and were then divided by the total time to get a single value for 'advection velocity' per bin. The advection velocities in Fig. 4m are normalized to the minimum value which coincides with the maximum value of the normalized medial MyoII intensity.

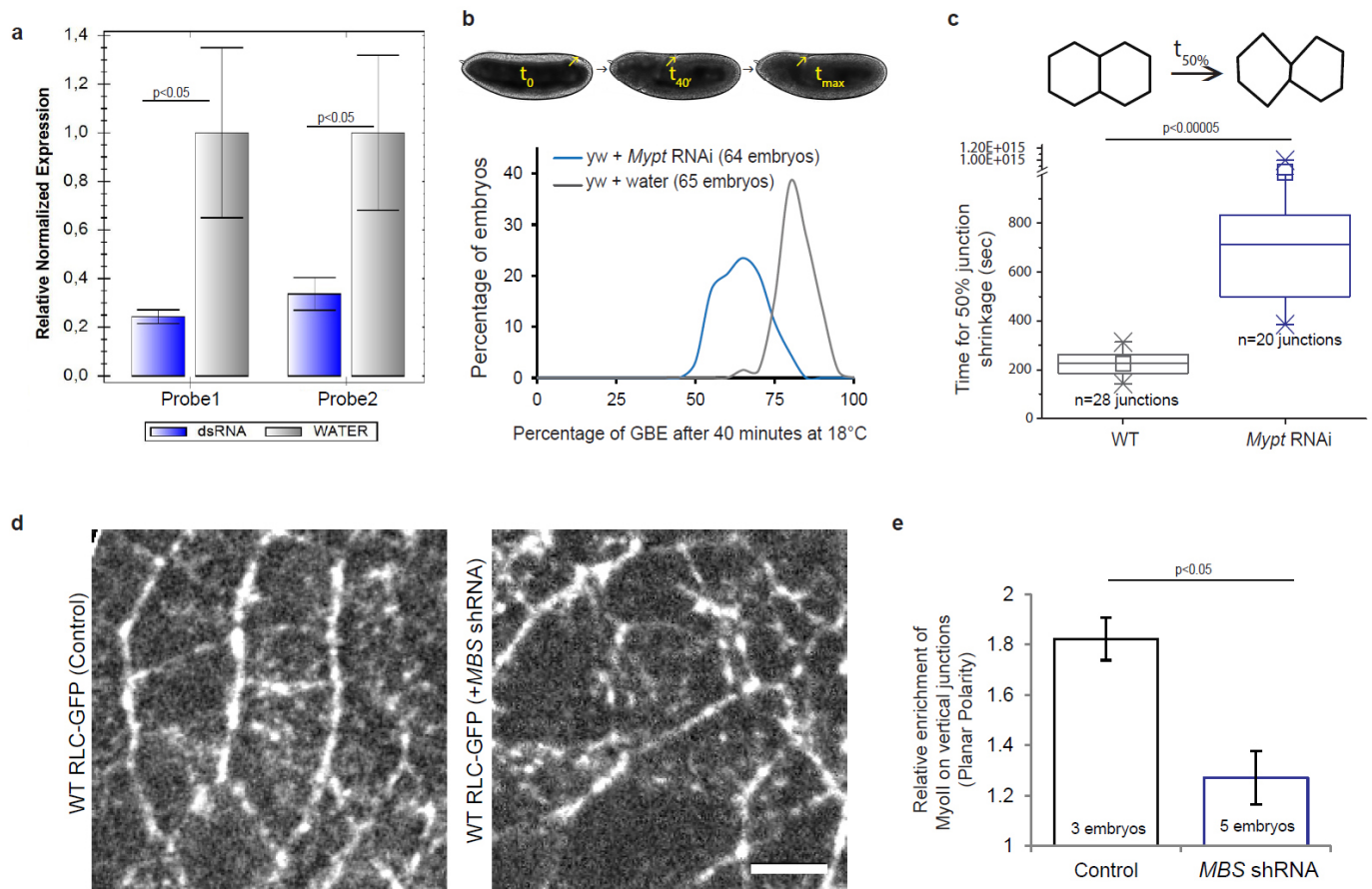
Statistics. Average values are calculated from ' n ', where n is the number of junctions/cells/pulses as noted in the figure. This number was used for conducting statistical significance tests. In planar polarity measurements (Fig. 1b, Extended Data Figs 1e and 4a), n is the number of embryos. Error bars usually depict s.e.m.

(that is, $\frac{s.d.}{\sqrt{n}}$, where n = number of embryos, unless otherwise specified). For

box plots, the open square indicates the mean, crosses indicate 99% or 1%, dashes are maximum or minimum and error bars are s.d. For large data sets ($n > 100$) such as junctional intensity measurements, two-sample Student's t -tests (two-tailed distribution with unequal variance) were used to determine P values. A t -test was performed only if the distribution was normal as tested with a Kolmogorov–Smirnov test. For small data sets ($n < 100$) and non-normal distributions, non-parametric Mann–Whitney U -tests were used to determine P values. For calculating the significance of difference of divergence values from 0, one-sample t -tests were used (null hypothesis: mean = 0). For comparing the distribution of cosine similarity indices with controls, two-independent-samples Kolmogorov–Smirnov tests were used. All statistical tests and box plots were made in OriginPro. No experiments were randomized and the investigator was not blinded to any data set. No statistical methods were used to predetermine sample size.

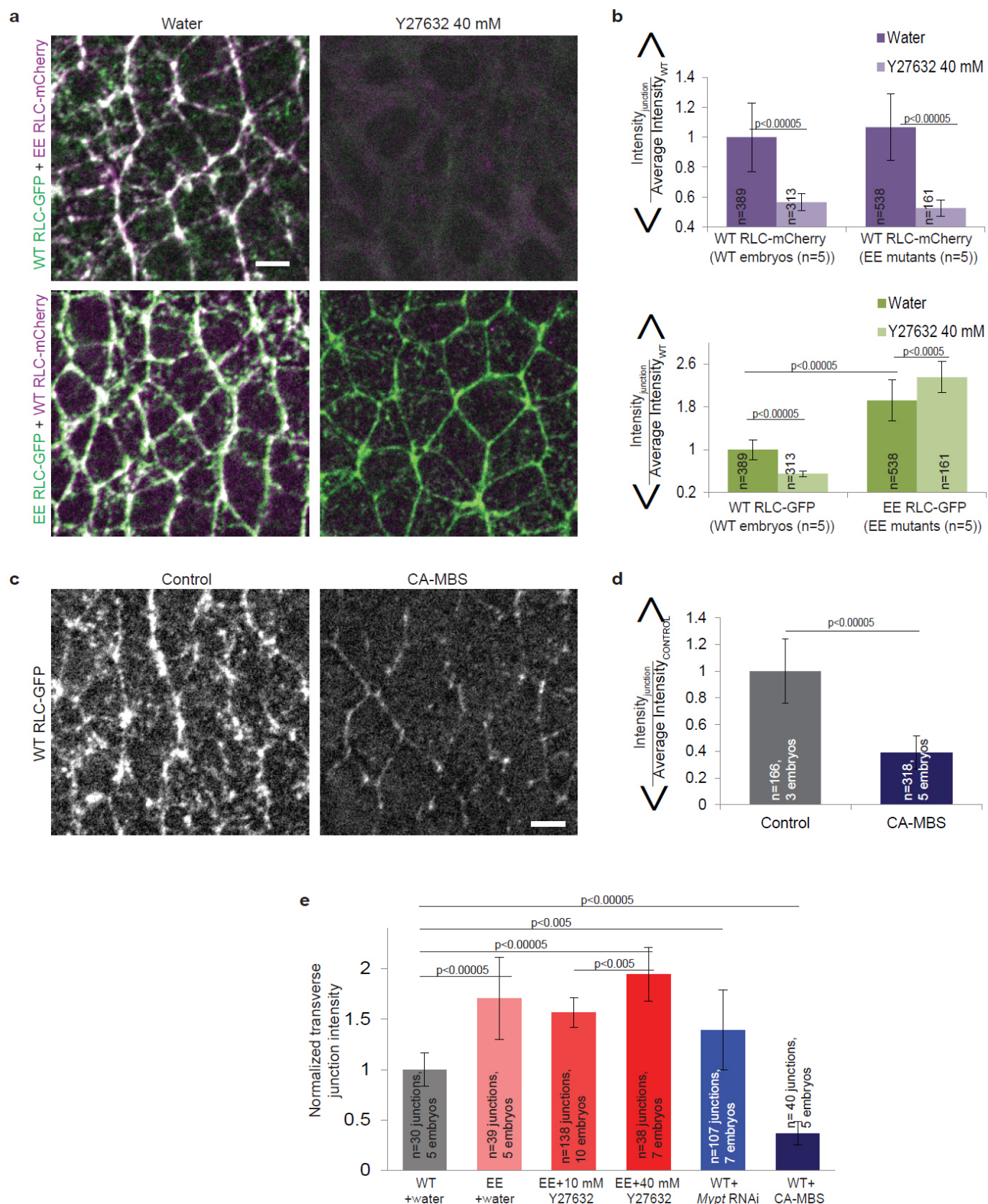
Repeatability. All measurements were made from 1–10 embryos from experiments replicated 2–10 times. Images and videos are representative of the data set.

30. Karess, R. E. *et al.* The regulatory light chain of nonmuscle myosin is encoded by *spaghetti-squash*, a gene required for cytokinesis in *Drosophila*. *Cell* **65**, 1177–1189 (1991).
31. Cavey, M. & Lecuit, T. Imaging cellular and molecular dynamics in live embryos using fluorescent proteins. *Methods Mol. Biol.* **420**, 219–238 (2008).



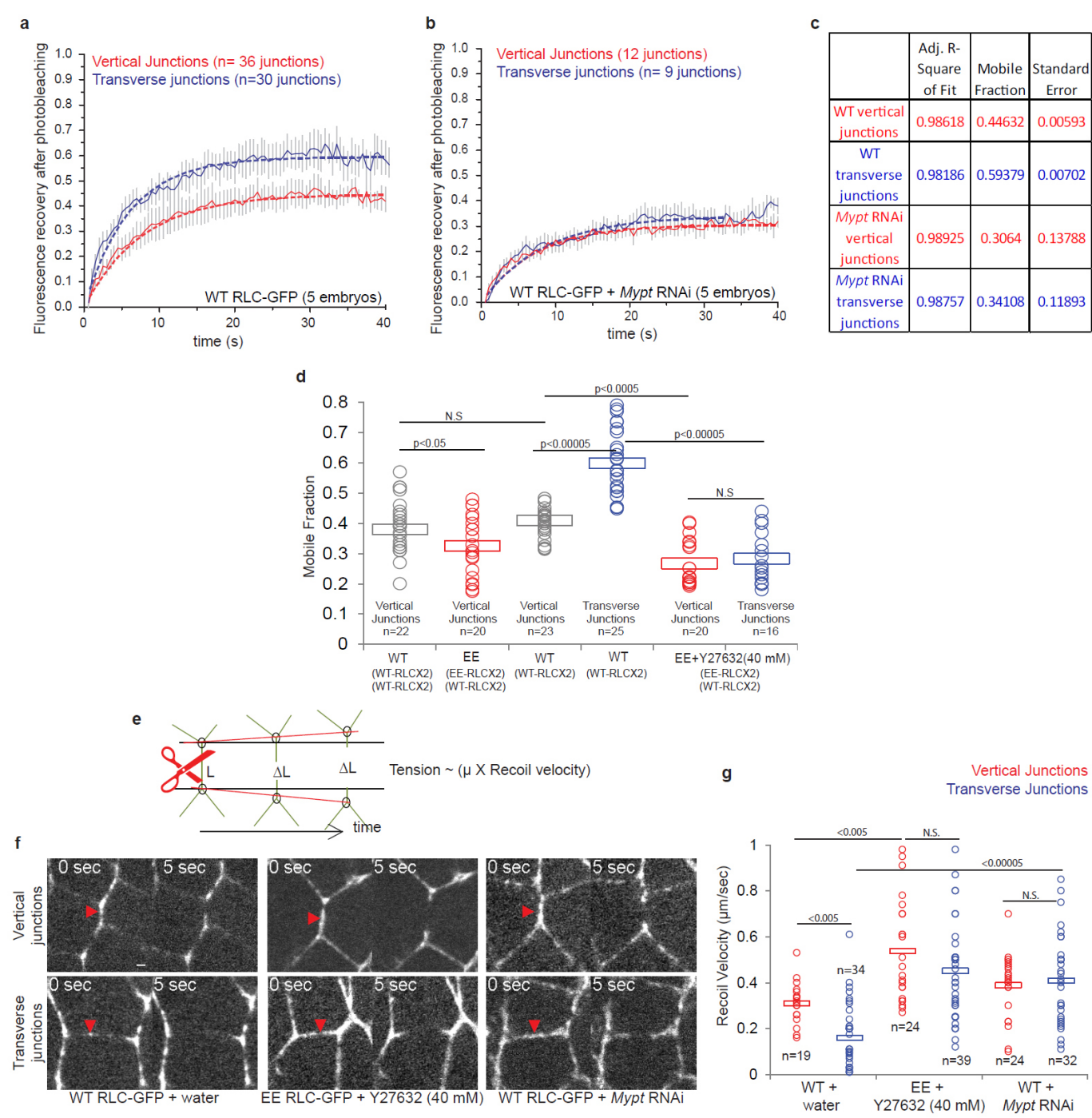
Extended Data Figure 1 | RNAi specifically downregulates MyoII phosphatase resulting in cellular and tissue-level defects. **a**, Gene expression data of *Mypt-75d* measured from quantitative PCR in water- and dsRNA-injected embryos using two different Taqman probes. P values calculated with unpaired t -test. **b**, Distribution of percentage of GBE in water and *Mypt*

dsRNA-injected embryos. **c**, Box plot of time taken for junction shrinkage in water and *Mypt* dsRNA-injected embryos. **d**, MyoII distribution in control (left) and MBS shRNA-expressing embryo (right). **e**, Average planar polarity of MyoII in these conditions. P values calculated with Mann-Whitney U -test. Scale bar, 5 μ m.



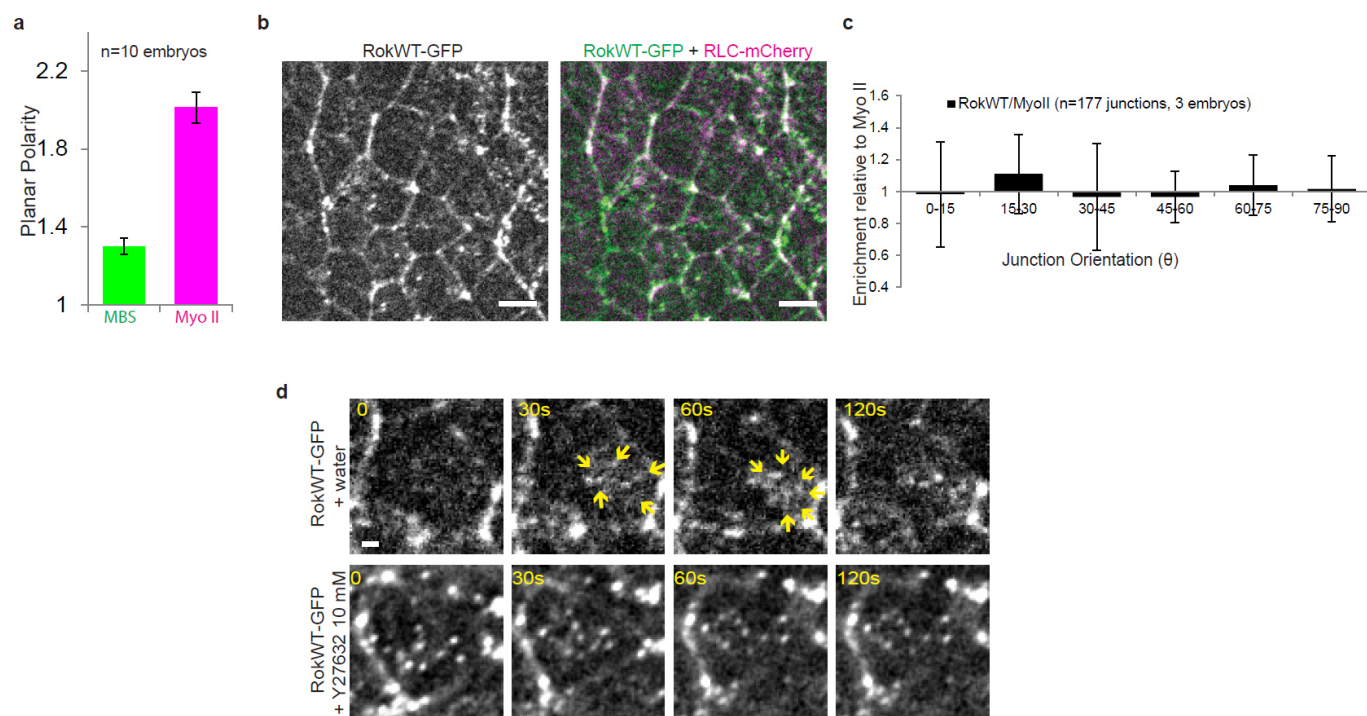
Extended Data Figure 2 | RLC phospho-cycling determines quantitative levels of MyoII at cell junctions. **a**, Distribution of MyoII in embryos co-expressing wild-type (WT) RLC-GFP with wild-type RLC-mCherry (top) and phosphomimetic RLC-GFP with wild-type RLC-mCherry (bottom) injected with water (left) or Y27632 (right). **b**, Mean intensity of MyoII in mCherry (top) and GFP (bottom) channels in these conditions. *P* values calculated with

Student's *t*-test. **c**, Distribution of MyoII in control embryos or in embryos expressing constitutively active MBS (CA-MBS). **d**, Average junctional intensity of MyoII in these conditions. *P* values calculated with Student's *t*-test. **e**, Average intensities of MyoII on transverse junctions in various conditions. *P* values calculated with Mann-Whitney *U*-test. Scale bar, 5 μ m.



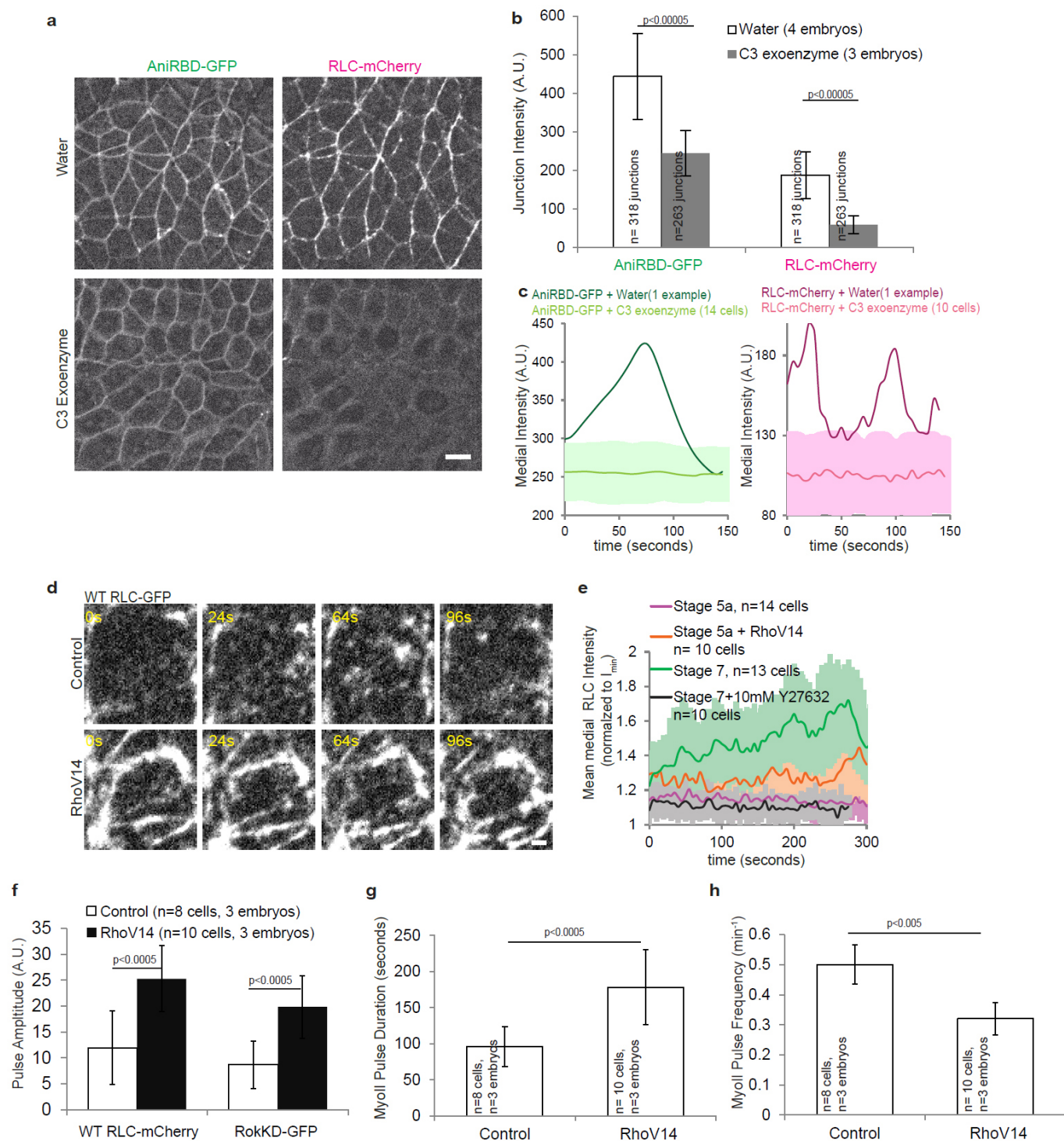
Extended Data Figure 3 | Proportions and cycling of phosphorylated RLC regulates MyoII turnover and mechanical tension. **a, b**, Mean fluorescence recovery time traces (normalized as described in methods) of RLC-GFP from vertical (red) and transverse junctions (blue) in wild-type (**a**) and *Mypt-75d* RNAi (**b**) embryos. The dotted lines are fits derived from the equation described in the methods. **c**, Table shows the goodness of fit (R^2) and the respective mobile fraction from fitted curves in **a** and **b**. **d**, Plot shows the mobile fraction of RLC in different conditions. Values from each junction

(open circles) and the average from all junctions (open rectangle) are shown. **e**, Cartoon showing scission of junctions and their elastic recoil in response **f**, Vertical and transverse junctions before (0 s) and after (5 s) laser ablation in different conditions. **g**, Recoil velocities of vertical and transverse junctions in different conditions. Values from each junction (open circles) and the average from all junctions (open rectangle) are shown. P values calculated with Mann-Whitney U -test. Scale bar, 1 μm .



Extended Data Figure 4 | Wild-type Rok is planar polarized and pulsatile during GBE. **a**, Average planar polarity of MBS and MyoII. **b**, Distribution of wild-type Rok (RokWT)-GFP without (left) and with (right) RLC-mCherry (top). Scale bar, 5 μ m. **c**, Average enrichment of RokWT-GFP relative to

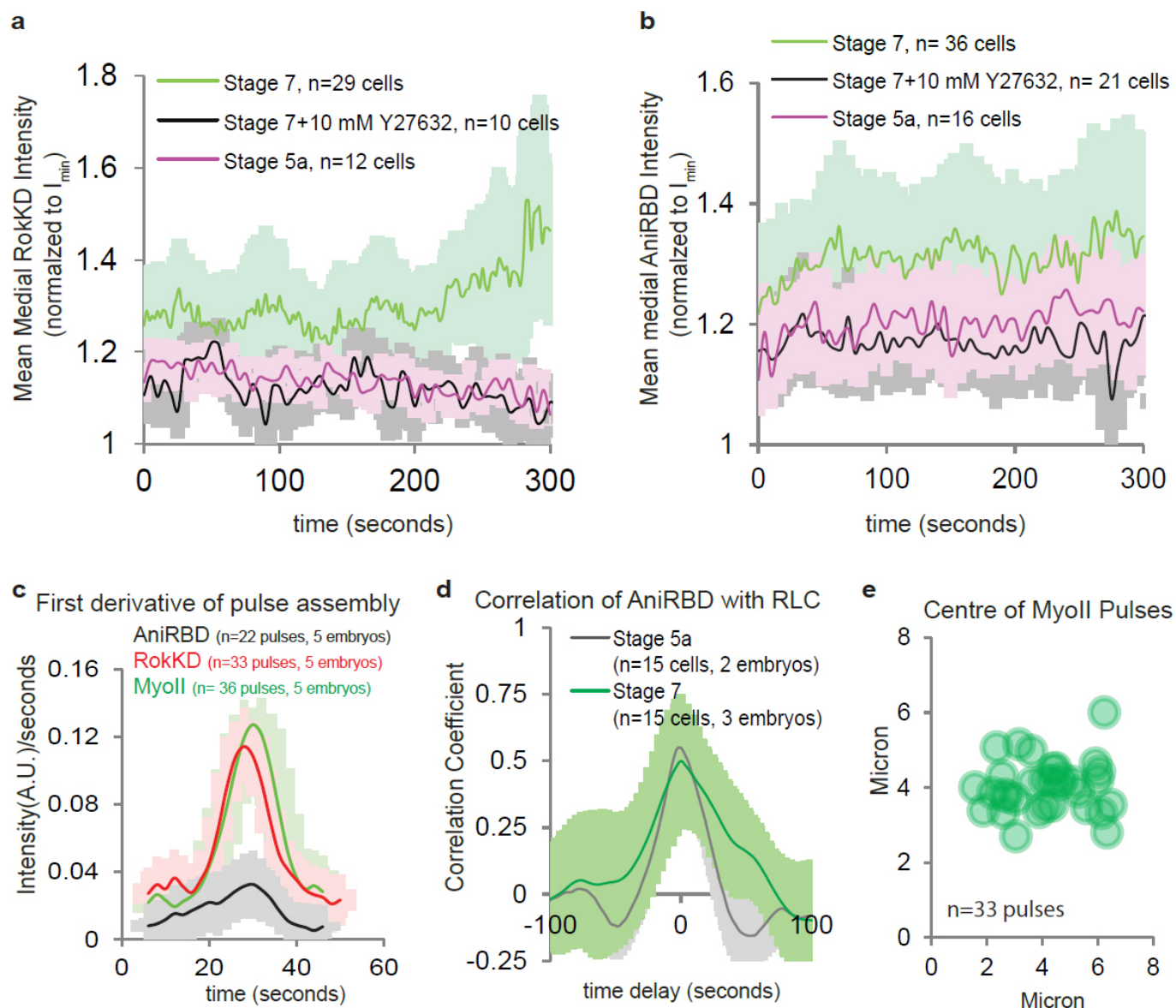
MyoII enrichment at junctions categorized by their angles (bottom). **d**, Time snapshots of RokWT-GFP embryos injected with water (top) or 10 mM Y27632 (bottom). Scale bar, 1 μ m.



Extended Data Figure 5 | Rho1 GTP is required for MyoII activation.

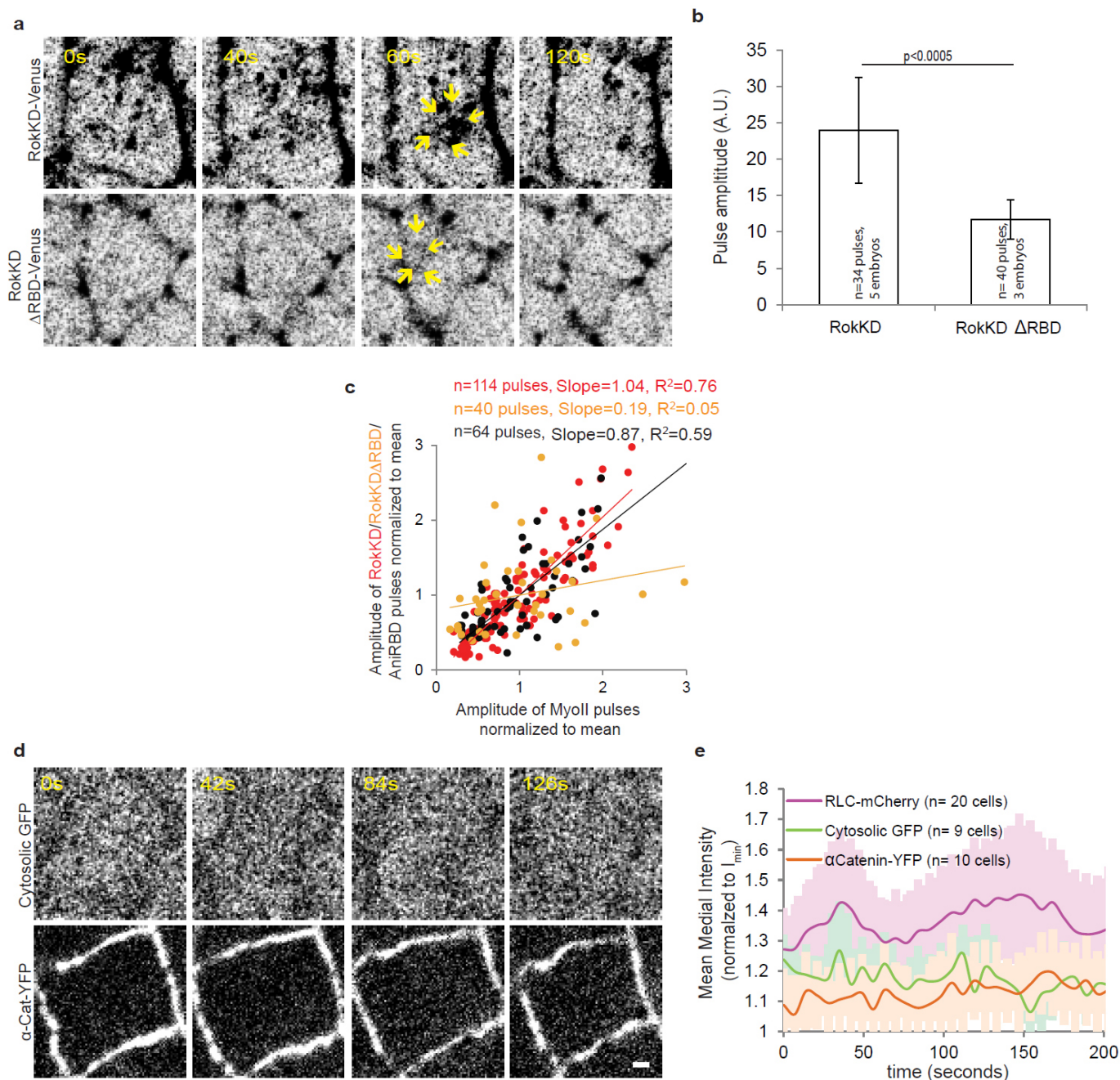
a, Representative images of AniRBD-GFP (left) and RLC-mCherry (right) in water (top) and C3 exoenzyme (bottom) injected embryos. Scale bar, 5 μm . **b**, AniRBD-GFP and RLC-mCherry mean junctional intensities in water and C3 exoenzyme injected embryos. *P* values calculated by Student's *t*-test. **c**, Mean medial intensity time trace of AniRBD-GFP (left) and RLC-mCherry (right) in C3 exoenzyme injected embryos compared to a representative water injected embryo. Error bars represent s.d. **d**, Time snapshots showing medial pool of

RLC-GFP in control and embryo expressing constitutively active Rho (RhoV14). Scale bar, 1 μm . **e**, Mean medial intensity time traces of RLC-GFP in different conditions and stages (Error bars: Standard deviation) **f**, Average pulse amplitudes of RLC-mCherry and kinase-dead Rok (RokKD)-GFP in control and RhoV14-expressing embryos **g**, **h**, Average duration and frequency of MyoII pulses in control and RhoV14-expressing embryos. *P* values calculated by Mann-Whitney *U* test.



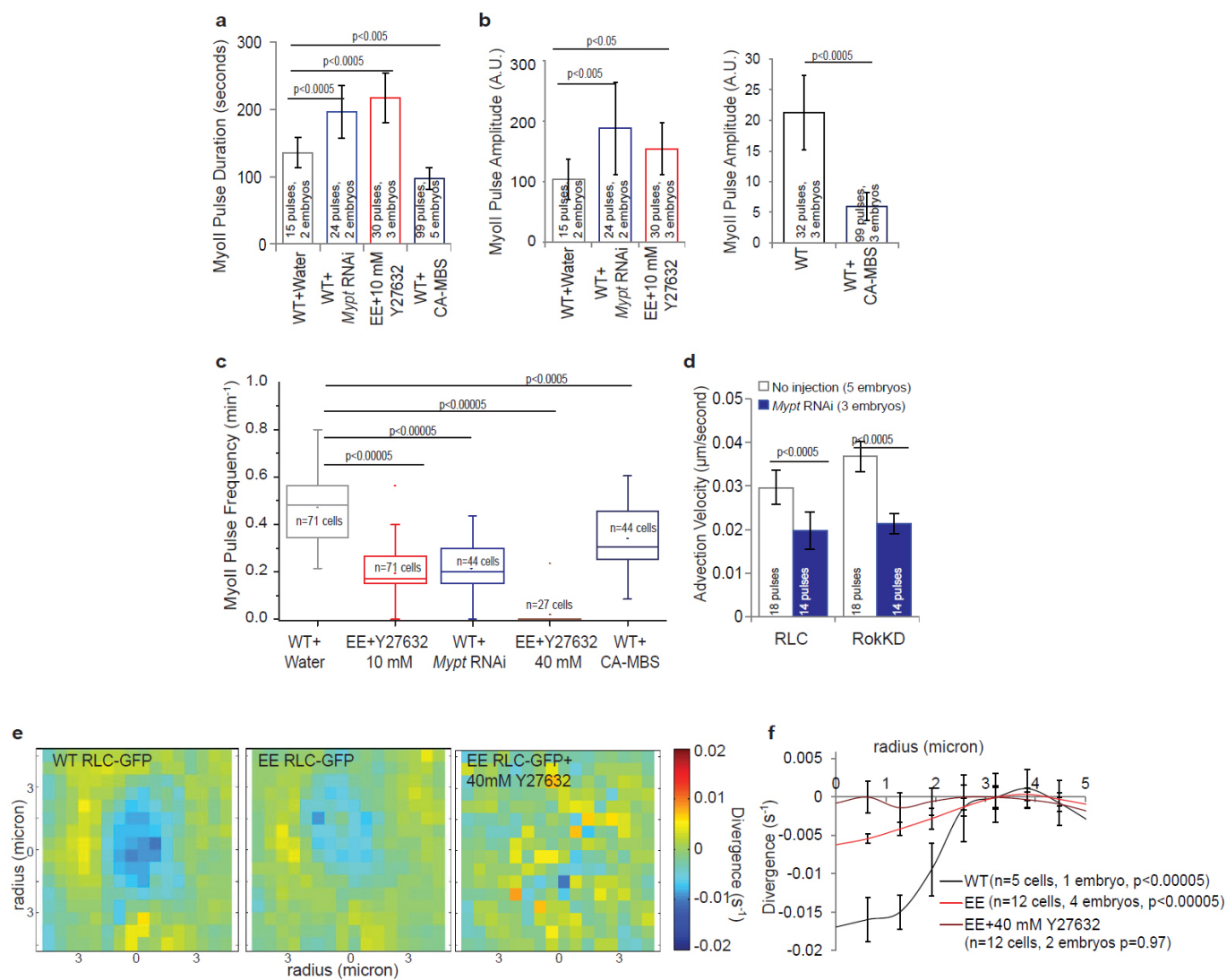
Extended Data Figure 6 | Non-linear MyoII accumulation from stochastic Rho1 activation. **a, b**, Mean medial intensity time traces of RokKD-GFP (**a**) and AniRBD-GFP (**b**) in various conditions. Error bars indicated s.d. **c**, Average rate of change (first derivative) of AniRBD-GFP, RokKD-GFP

and RLC-GFP medial intensities during pulse assembly. **d**, Temporal cross-correlation analysis of medial intensity time traces of RLC-mCherry with AniRBD-GFP during early GBE (stage 5a) and GBE (stage 7). **e**, Distribution of centres of masses of MyoII medial intensity at the time of pulse coalescence.



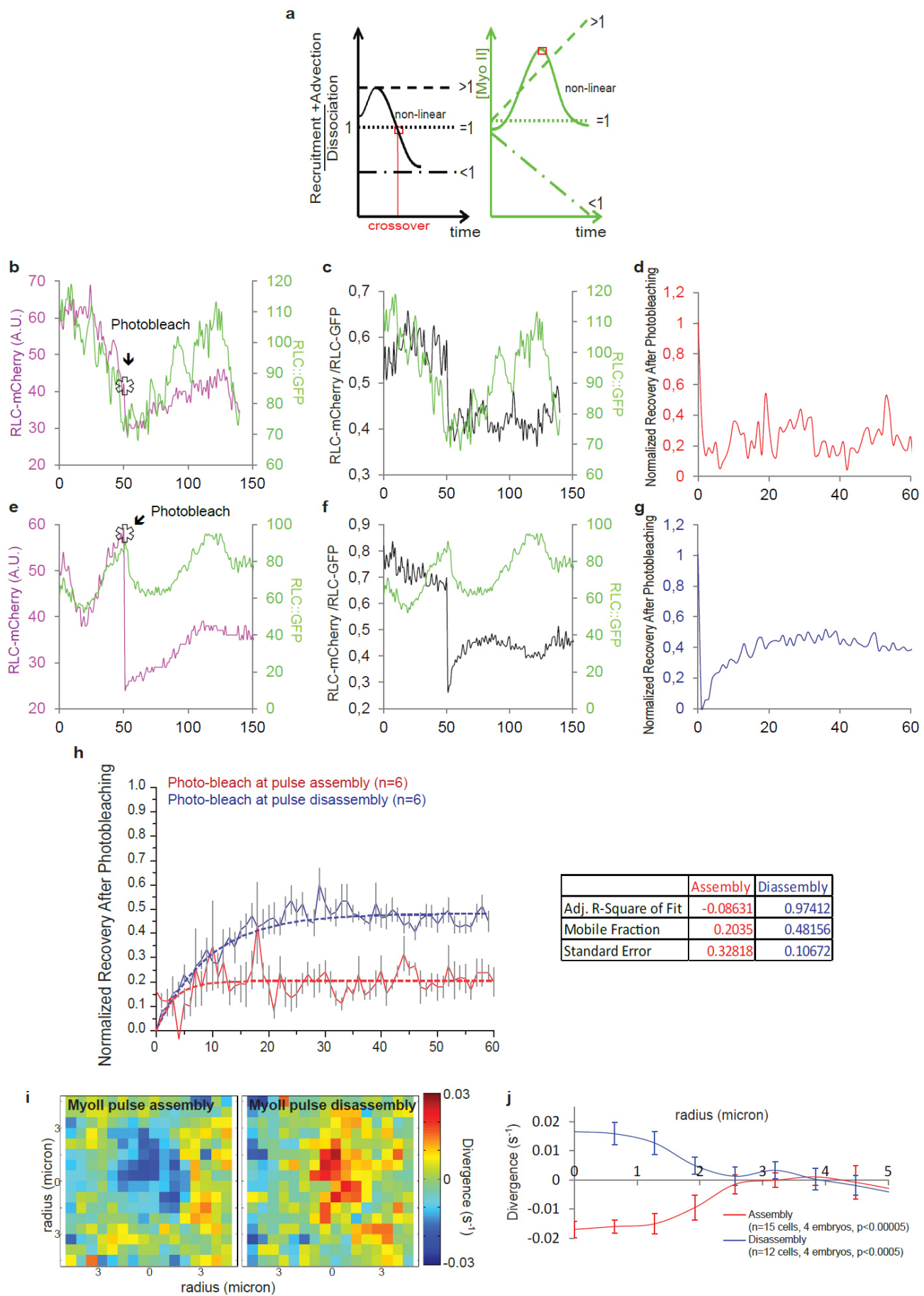
Extended Data Figure 7 | MyoII regulatory pathway components are advected. **a**, Time snapshot from medial pool of embryos expressing either RokKD-Venus (top) or RokKDΔRBD-Venus (bottom). **b**, Mean pulse amplitudes of RokKD-Venus and RokKDΔRBD-Venus. **c**, Pairwise correlation of RokKD-GFP, RokKDΔRBD-Venus and AniRBD-GFP pulse

amplitudes with the respective RLC-mCherry pulse amplitudes **d**, Time snapshot from medial pool of embryos expressing either cytosolic GFP (top) or α-catenin-YFP (bottom). **e**, Mean medial intensity time traces of RLC-mCherry, cytosolic GFP and α-catenin-YFP during GBE. Scale bar, 1 μm.



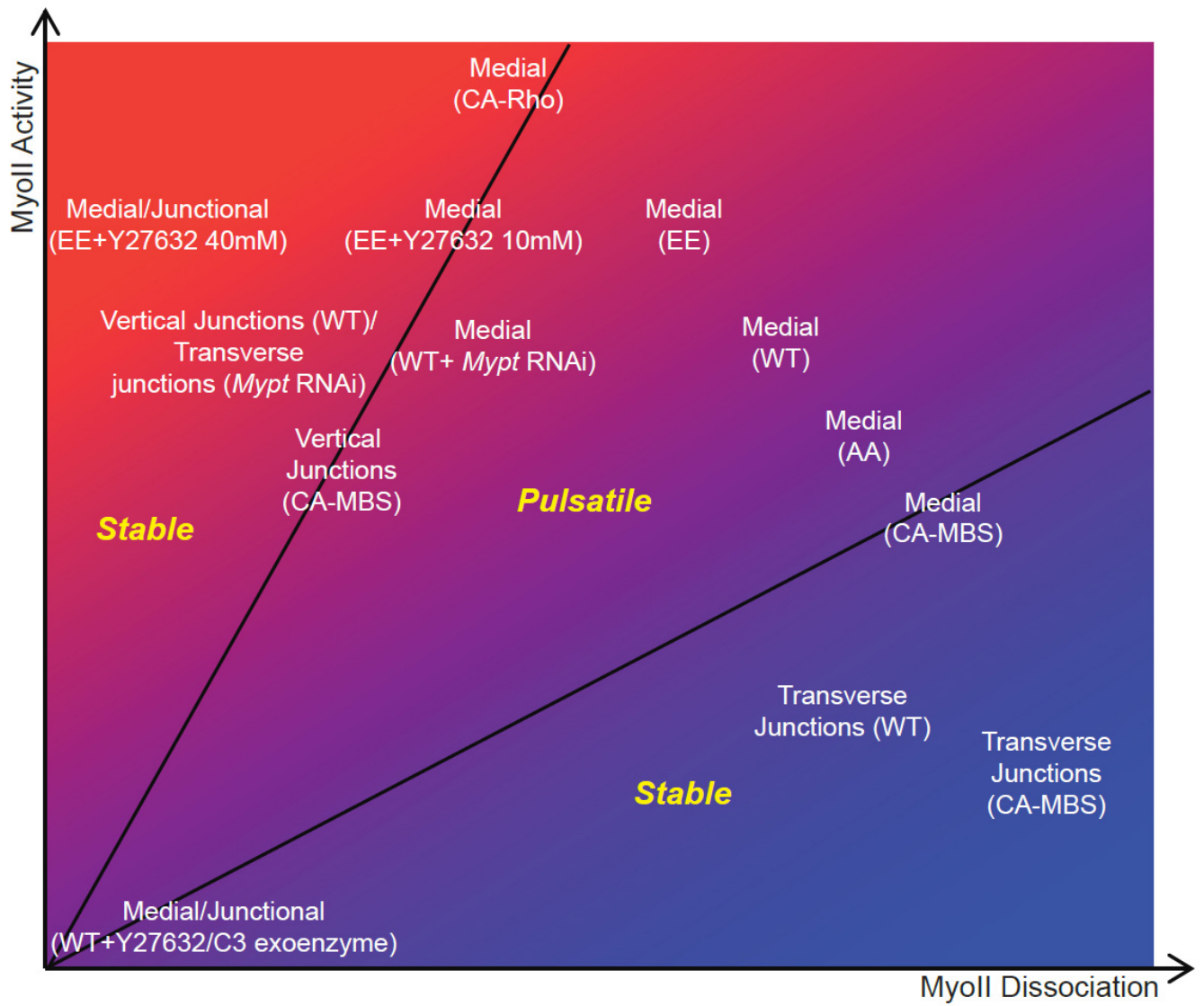
Extended Data Figure 8 | RLC dephosphorylation is required for pulse disassembly. **a–c**, Mean MyoII pulse duration (**a**), amplitude (**b**) and frequency (**c**) in various conditions. *P* values calculated with Mann–Whitney *U*-test. **d**, Average advection velocity of medial MyoII and RokKD in wild-type and *myoII* dsRNA-injected embryos. *P* values calculated with Mann–Whitney

U-test. **e**, Average divergence values from many cells of embryos expressing wild-type RLC–GFP (left), phosphomimetic (EE) RLC–GFP (middle) and phosphomimetic RLC–GFP injected with Y27632 (40 mM) (right). **f**, Plot showing average divergence values as a function of cell radius in these conditions. *P* values calculated with one-sample *t*-test.



Extended Data Figure 9 | Non-constant turnover and advection results in MyoII depletion during pulse disassembly. **a**, Behaviour of MyoII (green) for different ratios of recruitment + advection/dissociation (black) (dashed line, ratio >1 ; dotted line, ratio $=1$; dot-dash line, ratio <1 ; solid line: non-constant ratio starting from >1 and crossing over to <1). **b**, **e**, Medial intensity time traces of embryo co-expressing RLC-mCherry and RLC-GFP. The star shows the time of photo-bleach either at the onset of pulse assembly (**b**) or pulse disassembly (**e**). **c**, **f**, Time traces show the ratio of intensities from the red and the green channels (black). **d**, **g**, Fluorescence recovery curves obtained from

the red/green ratio. The curves are normalized as described in the methods. **h**, Average of fluorescence recovery curves from many different experiments. Error bars, s.e.m. where n = number of pulses. The dotted curve corresponds to the fit derived from the equation described in the methods. Table shows the goodness of fit (R^2) and the respective mobile fraction during pulse assembly and disassembly. **i**, Average divergence values of MyoII during pulse assembly (left) and disassembly (right). **j**, Plot shows average divergence values as a function of cell radius during pulse assembly and disassembly. P values calculated with one-sample t -test.



Extended Data Figure 10 | A qualitative phase diagram of MyoII dynamics. The phase diagram summarizes MyoII behaviour (stable or oscillatory) in the parameter space of MyoII dissociation and MyoII activity in different experimental conditions.

The zinc transporter ZIP12 regulates the pulmonary vascular response to chronic hypoxia

Lan Zhao^{1*}, Eduardo Oliver^{1*}, Klio Maratou⁴, Santosh S. Atanur⁴, Olivier D. Dubois¹, Emanuele Cotroneo¹, Chien-Nien Chen¹, Lei Wang¹, Cristina Arce¹, Pauline L. Chabosseau⁴, Joan Ponsa-Cobas³, Maria G. Frid⁷, Benjamin Moyon⁵, Zoe Webster⁵, Almaz Aldashev⁶, Jorge Ferrer³, Guy A. Rutter², Kurt R. Stenmark⁷, Timothy J. Aitman^{4†} & Martin R. Wilkins¹

The typical response of the adult mammalian pulmonary circulation to a low oxygen environment is vasoconstriction and structural remodelling of pulmonary arterioles, leading to chronic elevation of pulmonary artery pressure (pulmonary hypertension) and right ventricular hypertrophy. Some mammals, however, exhibit genetic resistance to hypoxia-induced pulmonary hypertension^{1–3}. We used a congenic breeding program and comparative genomics to exploit this variation in the rat and identified the gene *Slc39a12* as a major regulator of hypoxia-induced pulmonary vascular remodelling. *Slc39a12* encodes the zinc transporter ZIP12. Here we report that ZIP12 expression is increased in many cell types, including endothelial, smooth muscle and interstitial cells, in the remodelled pulmonary arterioles of rats, cows and humans susceptible to hypoxia-induced pulmonary hypertension. We show that ZIP12 expression in pulmonary vascular smooth muscle cells is hypoxia dependent and that targeted inhibition of ZIP12 inhibits the rise in intracellular labile zinc in hypoxia-exposed pulmonary vascular smooth muscle cells and their proliferation in culture. We demonstrate that genetic disruption of ZIP12 expression attenuates the development of pulmonary hypertension in rats housed in a hypoxic atmosphere. This new and unexpected insight into the fundamental role of a zinc transporter in mammalian pulmonary vascular homeostasis suggests a new drug target for the pharmacological management of pulmonary hypertension.

We have reported previously that the Fisher 344 (F344) rat strain is resistant to hypoxia-induced pulmonary hypertension compared with the Wistar Kyoto (WKY) strain². Linkage analysis of an F₂ population derived from inbred WKY × F344 rats identified a quantitative trait locus (QTL) on chromosome 17 (ref. 2). On the basis of this observation, we next conducted ten successive microsatellite-guided backcrosses of offspring with WKY rats and derived two congenic strains in which the original QTL was dissected and represented as partly overlapping regions of a donor F344 genome interposed onto the genetic background of the WKY recipient strain (Extended Data Figs 1 and 2). Resistance to hypoxia-induced pulmonary hypertension was detected in one of the congenic strains (R47A; Fig. 1a–d and Extended Data Fig. 2). Three subcongenic strains (SubA, SubB and SubC) were derived by further backcrosses of R47A onto the WKY background and the congenic interval was fine-mapped to a region of 8.28 megabase pairs (Mbp) containing an estimated 65 genes (rat chromosome 17: positions 85,072,475–93,347,784) (Fig. 1 and Extended Data Fig. 2a). Whole-genome sequencing (greater than 20× coverage) of the WKY and F344 parental strains⁴ revealed 13 non-synonymous coding SNPs affecting 9 genes within the refined congenic interval, and 6 indels

resulting in frameshift mutations in 4 genes (Extended Data Table 1). Polymorphic examination of the 13 SNPs and 6 indels in 2 additional rat strains susceptible to hypoxia-induced pulmonary hypertension (the spontaneously hypertensive and fawn-hooded rat strains, respectively) excluded 5 SNPs and 5 indels and narrowed the genes of interest to 7 (*Slc39a12*, *St8sia6*, *Cubn*, *Nmt2*, *Dclre1c*, *Hspa14* and *Cdnf*; Fig. 1e and Extended Data Table 1). Further polyphen analysis allowed us to exclude five listed genes (*St8sia6*, *Cubn*, *Nmt2*, *Dclre1c* and *Cdnf*) as the non-synonymous coding changes were predicted to be benign. We identified *Slc39a12*, with a loss of thymidine at position 88,575,534 leading to a frameshift mutation in exon 11, as the highest priority candidate gene for further investigation.

Slc39a12 encodes the solute carrier 39 zinc transporter family (ZIP1–14) member 12 (ZIP12) and has high specificity for zinc⁵. The ZIP family tightly regulates cellular zinc homeostasis in numerous cell types by promoting zinc uptake from the extracellular space or release from intracellular compartments. The rat *Slc39a12* gene contains 12 exons and the ZIP12 protein comprises 688 amino acids with a secondary structure comprising 8 transmembrane domains. In the F344 strain, the frameshift mutation in *Slc39a12* introduces a stop-codon predicting a carboxy (C)-terminal truncated ZIP12 protein of 553 amino acids (Extended Data Fig. 3a). This affects the conserved zinc transporting aqueous cavity between transmembrane domains IV and V, resulting in the loss of the metalloprotease motif (HEXPHE), which would be expected to lead to a reduction in zinc transport⁶.

A pathognomonic histological signature of chronic hypoxia-induced pulmonary hypertension is thickening of the pulmonary vascular media (owing to hyperplasia and hypertrophy of smooth muscle cells) and the muscularization of previously unmuscularized pulmonary arterioles⁷. We found that ZIP12 messenger RNA (mRNA) levels were very low and ZIP12 protein was undetectable by immunohistochemistry in the pulmonary vasculature of adult WKY rats housed in a normal oxygen atmosphere, but WKY rats exposed to hypoxia showed markedly increased lung ZIP12 mRNA levels and pronounced ZIP12 expression in remodelled pulmonary arterioles (Fig. 2a, b and Extended Data Fig. 3b). ZIP12 expression was evident in vascular smooth muscle but also other cell types (endothelial and interstitial cells) known to contribute to structural changes seen in hypoxic lungs. In contrast, and consistent with a frameshift mutation in *Slc39a12* predicting a C-terminal truncated protein, ZIP12 was undetectable with an antibody directed at the C terminus of the protein in the lungs of chronically hypoxic F344 rats (Fig. 2b and Extended Data Fig. 3b).

Slc39a12 is highly conserved across species⁸ and transcribed constitutively in many tissues (<http://www.biogps.org>). To investigate

¹Centre for Pharmacology and Therapeutics, Division of Experimental Medicine, Imperial College London, Hammersmith Hospital, London W12 0NN, UK. ²Section of Cell Biology and Functional Genomics, Division of Diabetes, Endocrinology and Metabolism, Imperial College London, Hammersmith Hospital, London W12 0NN, UK. ³Section of Epigenomics and Disease, Department of Medicine, Faculty of Medicine, Imperial College London, Hammersmith Hospital, London W12 0NN, UK. ⁴Physiological Genomics and Medicine Group, Medical Research Council Clinical Sciences Centre, Hammersmith Hospital, London W12 0NN, UK. ⁵Transgenics and Embryonic Stem Cell Laboratory, Medical Research Council Clinical Sciences Centre, Hammersmith Hospital, London W12 0NN, UK. ⁶Institute of Molecular Biology and Medicine, 3 Togolok Moldo Street, Bishkek 720040, Kyrgyzstan. ⁷Department of Pediatrics and Medicine, Division of Critical Care Medicine and Cardiovascular Pulmonary Research Laboratories, University of Colorado Denver, Denver, Colorado 80045, USA. [†]Present address: Centre for Genomic and Experimental Medicine, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.

*These authors contributed equally to this work.

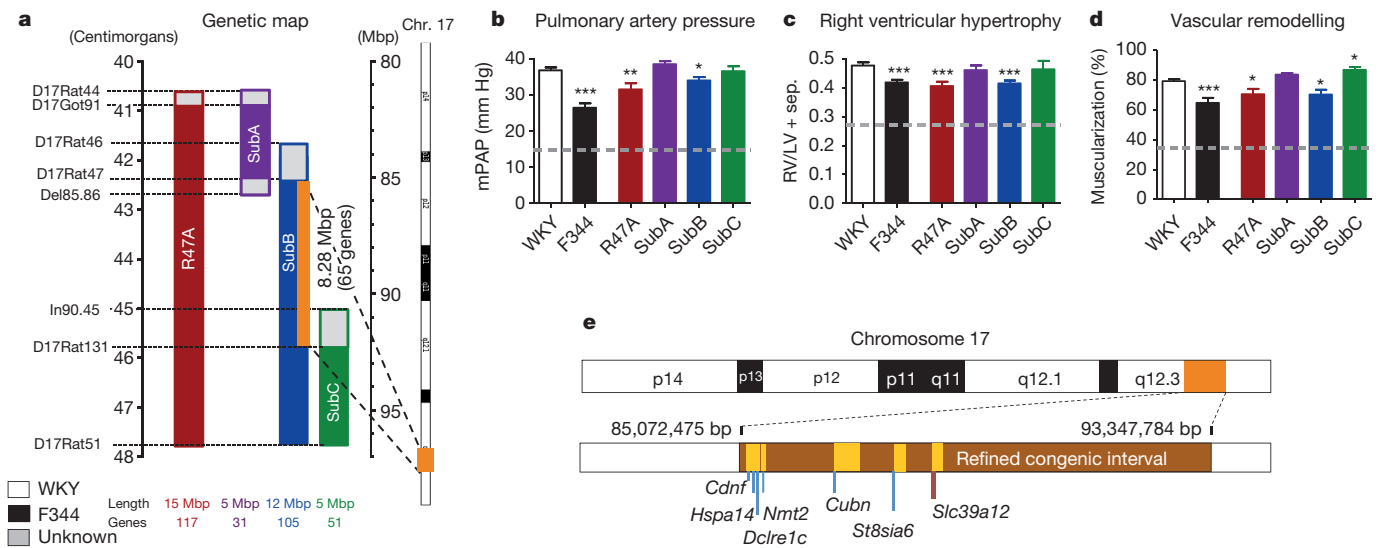


Figure 1 | The pulmonary vascular response to hypoxia in the F344 rat is influenced by a region of chromosome 17 containing *Slc39a12*. **a**, A genetic map of three sub-congenic strains (SubA, SubB and SubC) derived from the R47A congenic strain (originally derived from a WKY × F344 cross) backcrossed with the WKY parental strain. The refined congenic region (orange) of 8.28 Mbp containing 65 genes is within the SubB strain. **b–d**, SubB exhibits attenuated pulmonary hypertension after 2 weeks exposure to a 10% O₂ atmosphere compared with WKY, SubA and SubC rats: **b**, mean pulmonary artery pressure (mPAP); **c**, right ventricular hypertrophy (ratio of mass of right ventricle to left ventricle plus septum, RV/LV + sep.) ($n = 17$ WKY, 15 F344,

14 R47A, 8 SubA, 10 SubB, 10 SubC); **d**, vascular muscularization ($n = 6$ per group). Dotted line indicates mean measurements from all the rats in a normal oxygen atmosphere (21% O₂; mPAP = 14.7 ± 0.3 mm Hg; right ventricular hypertrophy = 0.270 ± 0.004 ; percentage muscularization = 34.2 ± 0.36 ; for actual values in rat strains see Extended Data Fig. 2c–e). Values are expressed as mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared with WKY after one-way analysis of variance (ANOVA) followed by Bonferroni correction for multiple testing. **e**, The genes of interest (*Slc39a12*, *St8sia6*, *Cubn*, *Nmt2*, *Dclre1c*, *Hspa14* and *Cdnf*) identified within the SubB congenic interval. The frameshift mutation in *Slc39a12* introduces a stop-codon, resulting in a truncated protein.

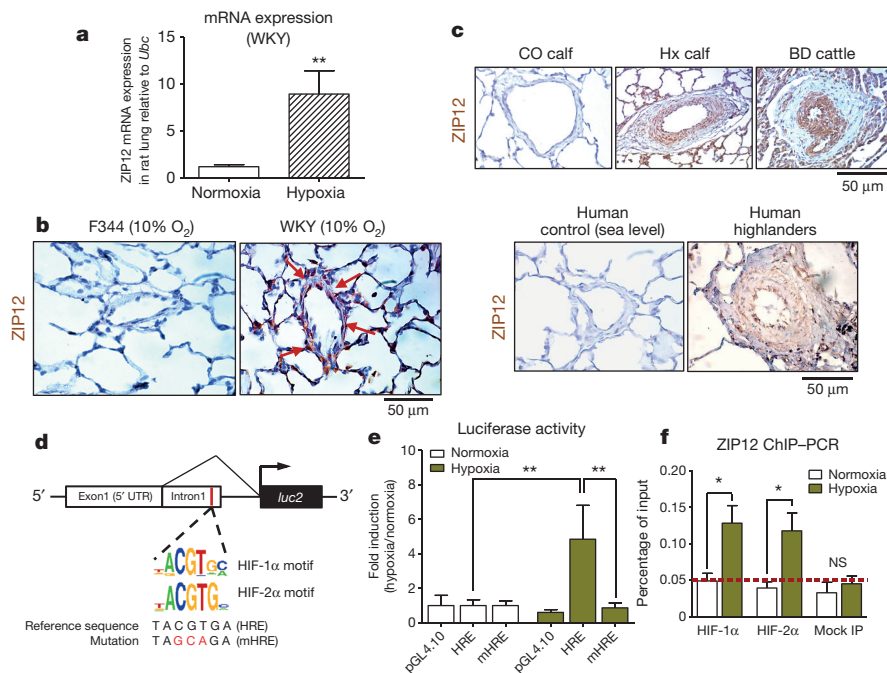


Figure 2 | *Slc39a12* encodes a zinc transporter, ZIP12, which is upregulated in pulmonary vascular tissue from mammals exposed to chronic hypoxia. **a**, ZIP12 mRNA levels in control and hypoxic WKY rat lungs ($n = 6$ per group). **b**, Prominent ZIP12 immunostaining in remodelled pulmonary arterioles in WKY (red arrows) but not F344 rat lungs exposed to hypoxia. **c**, No ZIP12 staining was detected in pulmonary arteries of low-altitude (normoxia control, CO calf) calves and sea-level humans, yet prominent ZIP12 immunostaining was observed in the remodelled pulmonary arteries of calves with severe pulmonary hypertension (Hx calf), in cattle naturally susceptible to pulmonary hypertension at altitude ('Brisket disease', BD) as well as Kyrgyz highlanders residing above 2,500 m. **d**, Design of the luciferase reporter vector pGL4.10 containing a 5' region of ZIP12 that includes an HRE encoding for both HIF-1α- and HIF-2α-binding motifs or a mutant HRE sequence where the

5'-ACGTG-3' motif has been replaced by 5'-AGCAG-3' (mHRE). **e**, Human pulmonary artery smooth muscle cells (HPASMCs) transfected with the ZIP12 HRE reporter vector demonstrated significantly increased luciferase activity after exposure to hypoxia, but not in the cells transfected with the mutant HRE vector ($n = 6$ per group, replicated twice). **f**, Increased levels of HIF-1α and HIF-2α bound to the ZIP12 HRE assayed by ChIP-qPCR of chromatin from HPASMCs cultured in normoxia and hypoxic conditions ($n = 3$ per group, replicated twice). Data are calculated as the percentage of input levels, with the dotted line marking percentages below mock immunoprecipitation (IP). Values are expressed as mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared with normoxic control after one-way ANOVA followed by Bonferroni correction for multiple testing. NS, not significant.

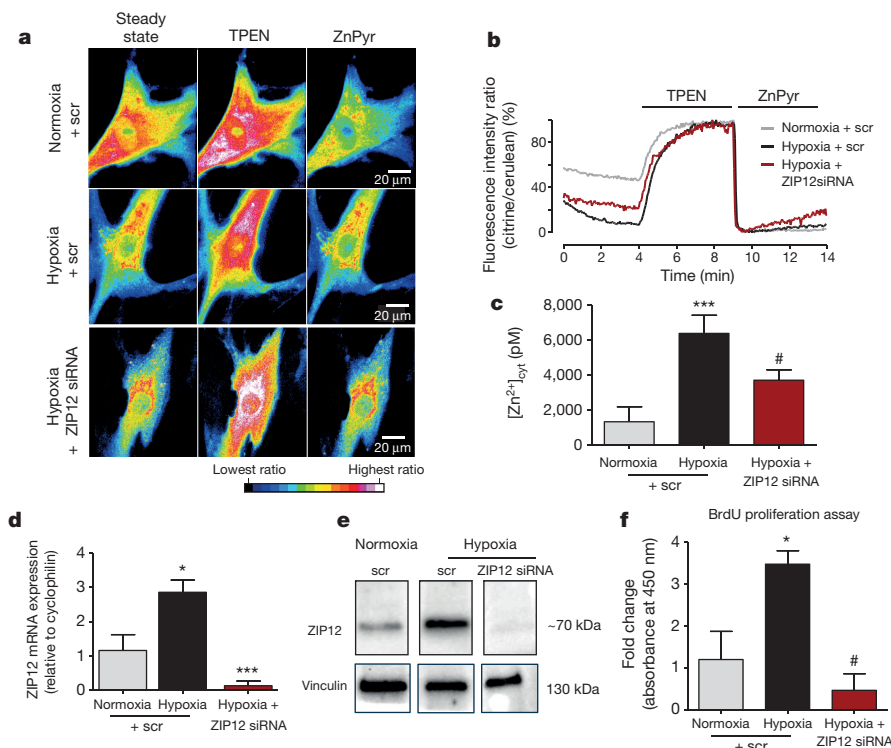


Figure 3 | ZIP12 knockdown inhibits hypoxia-induced increase in intracellular labile zinc concentration and proliferation of HPASMCs.

a, Representative wide-field microscope images of HPASMCs transfected with eCALWY-4 probe. Hypoxia exposure produced a striking increase in intracellular free zinc (resulting in decreased FRET)¹⁰. This was inhibited by transfection with ZIP12 siRNA. TPEN-mediated Zn^{2+} chelation was used to derive maximum fluorescence and 100 μ M $ZnCl_2$ in the presence of the Zn^{2+} ionophore and pyrithione (ZnPyr) was used to derive the minimum fluorescence. **b**, Representative traces showing the changes in fluorescence ratio of the eCALWY-4 probe. Steady-state fluorescence intensity ratio of citrine/ cerulean (R) was measured, then maximum and minimum ratios were determined to calculate free Zn^{2+} concentration using the formula $[Zn^{2+}] = K_d \times (R_{max} - R)/(R - R_{min})$, where the dissociation constant (K_d)

for eCALWY-4 is 630 pM, the maximum ratio (R_{max}) was obtained upon intracellular zinc chelation with 50 μ M TPEN and the minimum ratio (R_{min}) was obtained upon zinc saturation with 100 μ M $ZnCl_2$ in the presence of the Zn^{2+} ionophore pyrithione (5 μ M) (ref. 10). **c**, Quantification of intracellular zinc levels ($n = 10$ each group). **d**, Chronic hypoxia (48 h) increases ZIP12 mRNA levels in HPASMCs, which is inhibited by *Slc39a12* siRNA ($n = 5$ per group). **e**, Representative immunoblot of ZIP12 demonstrating inhibition of hypoxia-stimulated ZIP12 protein expression by *Slc39a12* siRNA in HPASMCs ($n = 3$). **f**, ZIP12 siRNA inhibits hypoxia-induced proliferation in HPASMCs ($n = 5$ per group, replicated twice). * $P < 0.05$, *** $P < 0.001$ compared with control group, # $P < 0.05$ compared with hypoxia group; scr, scrambled siRNA control.

the relevance of our observations in rats to other susceptible animal species, as well as humans, we examined ZIP12 expression in whole lung samples of (1) neonatal calves housed in a normal atmosphere or exposed to hypobaric hypoxia for 2 weeks (barometric pressure = 445 mm Hg (1 mm Hg = 133.3 Pa), equivalent to 4,500 m altitude, 12% O_2), (2) older (yearling) cattle with naturally occurring pulmonary hypertension (so-called 'Brisket disease') developed as a result of prolonged pasturing at high altitude (2,438–3,505 m) and (3) human subjects at sea level and Kyrgyz highlanders residing above 2,500 m. ZIP12 expression, which is undetectable by immunohistochemistry in healthy bovine and human lung exposed to a normal oxygen atmosphere (Fig. 2c), was clearly visible in the remodelled pulmonary vessels from chronic hypoxia exposure, indicating that ZIP12 upregulation in pulmonary vasculature is a common response to hypoxia (Fig. 2c).

To better understand the regulation of ZIP12 by hypoxia, we exposed human pulmonary vascular smooth muscle cells in culture to hypoxia (2% O_2). Increased hypoxia-inducible factor (HIF) protein and ZIP12 gene expression was observed in hypoxic cells; mRNA levels of other zinc transporters, ZIP6, ZIP7, ZIP10 and ZnT8, were unchanged (Extended Data Fig. 4a, b). Further examination of the *Slc39a12* gene using HOMER analysis⁹ revealed a hypoxia response element (HRE) encoding both HIF-1 α - and HIF-2 α -binding motifs (Fig. 2d) located 1 kb downstream of the ZIP12 transcription start site (human (hg19) chr10: 18,241,879–18,241,887). We cloned a 1.5 kb fragment of the 5' region of ZIP12 containing this HRE into the

luciferase reporter vector, pGL4.10 (Fig. 2d). Human pulmonary vascular smooth muscle cells transfected with the ZIP12 HRE reporter vector demonstrated significantly increased luciferase activity after exposure to hypoxia, while the luciferase activity of cells transfected with the mutant HRE vector (a substitution of the 5'-ACGTG-3' motif by 5'-AGCAG-3'; Fig. 2d) remained at basal normoxia levels (Fig. 2e). Chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR) confirmed the enrichment of both HIF-1 α and HIF-2 α bound to this ZIP12 HRE after hypoxia exposure (Fig. 2f).

We then explored the contribution of ZIP12 to the regulation of intracellular zinc levels in human pulmonary vascular smooth muscle cells. Intracellular labile zinc measured using a genetically encoded fluorescence resonance energy transfer (FRET)-based zinc probe, eCALWY-4 (ref. 10), exhibited a striking increase in cells exposed to hypoxia for 48 h, which was markedly reduced by inhibiting ZIP12 expression with a targeted short interfering (si)RNA (Fig. 3a–e). Inhibition of ZIP12 expression with siRNA also inhibited hypoxia-induced pulmonary vascular smooth muscle cell proliferation (Fig. 3f). ZIP12 siRNA transfection did not affect intracellular zinc levels or proliferation in normoxia (Extended Data Fig. 4c–f). These data suggest that disrupted ZIP12 expression exerts a direct effect on pulmonary vascular cells in response to hypoxia and contributes to the resistant pulmonary hypertension phenotype exhibited in F344 strain.

To provide direct genetic confirmation that disrupted ZIP12 expression attenuates the pulmonary vascular response to hypoxia, we employed zinc finger nuclease technology¹¹ to introduce mutations

in *Slc39a12* in the hypoxia-susceptible WKY rat strain. A mutant line was generated containing a frame-shift resulting in a truncated ZIP12 protein with loss-of-function (Extended Data Fig. 5). Inter-cross of heterozygous animals generated homozygous ($ZIP12^{-/-}$), heterozygous ($ZIP12^{+/-}$) and wild-type (WT) rats that were then exposed to hypoxia (10% O_2) for 2 weeks. $ZIP12^{-/-}$ rats demonstrated lower pulmonary artery pressures, reduced right ventricular hypertrophy and less vascular remodelling than WT rats (Fig. 4a–c and Extended Data Fig. 6a–d) with $ZIP12^{+/-}$ rats exhibiting an intermediate phenotype. WT rats resembled WKY rats after exposure to hypoxia, showing markedly increased lung ZIP12 expression in the remodelled pulmonary arterioles, in contrast to the absence of expression in $ZIP12^{-/-}$ rats (Fig. 4d, e). Comparison of the $ZIP12^{-/-}$ response to hypoxia with the WKY and F344 parental strains revealed that mutation of *Slc39a12* was responsible for about 50% of the resistance observed in the F344 strain, highlighting the importance of *Slc39a12* as a hypoxia-susceptibility gene but also suggesting that other genes yet to be identified might also contribute.

Systemic blood pressure and cardiac output in the hypoxic $ZIP12^{-/-}$ rats was similar to that of WT rats (Extended Data Fig. 6e–g), signifying that the reduced pulmonary artery pressures in the $ZIP12^{-/-}$ rat in chronic hypoxia is due to reduced pulmonary vascular resistance (PVR; mean pulmonary artery pressure = PVR \times cardiac output). Both vascular tone and structural remodelling contribute to PVR, and increased pulmonary vascular tone precedes the structural changes. ZIP12 expression may increase PVR by increasing pulmonary vascular tone. Zinc thiolate signalling has been reported to mediate the constriction of pulmonary microvascular endothelial cells in acute hypoxia through activation of protein kinase C and inhibition of myosin light chain phosphatase, inducing stress fibre formation and endothelial cell contraction¹². We have shown that ZIP12-targeted siRNA attenuates stress fibre formation in human pulmonary vascular smooth muscle cells cultured

in hypoxia (Extended Data Fig. 6h, i). However, given the time-dependent induction of ZIP12 expression in pulmonary vasculature by hypoxia, the main contribution of ZIP12 is likely to be in regulating the response to chronic rather than acute hypoxia. In further support of a direct effect on structural remodelling of pulmonary arterioles, we investigated angiogenesis *ex vivo* using pulmonary arteriole rings dissected from $ZIP12^{-/-}$ and WT rats. Vascular outgrowth from $ZIP12^{-/-}$ vessels in response to hypoxia was attenuated compared with vessels from WT rats (Extended Data Fig. 6j, k).

The underlying mechanisms through which ZIP12 affects hypoxic responses remain to be defined. Excess intracellular zinc concentrations mediated by upregulation of ZIP family members have been observed in a variety of tumour tissues and linked to cell proliferation and survival^{13–15}. Zinc is a structural component of a large variety of intracellular proteins, including enzymes and transcription factors. Zinc-binding motifs have been identified in drug targets for pulmonary hypertension, for example phosphodiesterase type 5 (PDE5) and histone deacetylases^{16,17}. Reduced ZIP12 expression and intracellular labile zinc levels would be expected to inhibit PDE5 activity¹⁸, and we have previously shown that PDE5 inhibition attenuated pulmonary vascular smooth muscle proliferation in culture¹⁹.

Following on from our demonstration that ZIP12 is hypoxia inducible and a key regulator of the pulmonary vascular response to chronic alveolar hypoxia exposure, we examined lung ZIP12 expression in other presentations of pulmonary hypertension where tissue hypoxia is an important driver of pathology. Again, in contrast to healthy lungs, ZIP12 expression was clearly evident in lung tissues from chronic iron-deficient rats²⁰ and rats exposed to monocrotaline, as well as in patients with idiopathic pulmonary arterial hypertension (IPAH)²¹ (Fig. 4f), prominent in the remodelled pulmonary vasculature as identified by co-staining with smooth muscle actin (Fig. 4g). HIF activation in these tissues was confirmed by upregulation of carbonic anhydrase IX, a recognized HIF-regulated biochemical signature

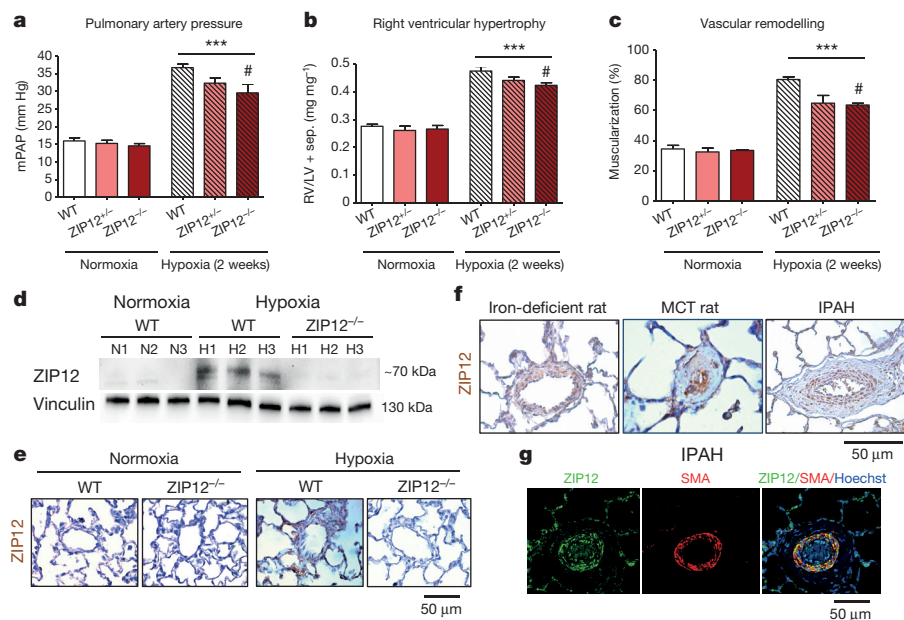


Figure 4 | Genetic disruption of ZIP12 in WKY rat attenuates hypoxia-induced pulmonary hypertension. **a–c**, Zinc finger nucleases were used to disrupt ZIP12 in the WKY strain. Rats deficient in ZIP12 demonstrate allele dose-dependent attenuation of hypoxia-induced pulmonary hypertension compared with WT rats: **a**, mPAP; **b**, right ventricular hypertrophy (normoxia groups: $n = 10$ WT, 8 $ZIP12^{+/-}$, 12 $ZIP12^{-/-}$; hypoxia groups: $n = 14$ WT, 16 $ZIP12^{+/-}$, 12 $ZIP12^{-/-}$); **c**, pulmonary arteriole muscularization ($n = 5$ per group). *** $P < 0.001$ compared with normoxia WT group, # $P < 0.05$ compared with hypoxia WT group after one-way ANOVA followed by

Bonferroni correction for multiple testing. **d**, ZIP12 was undetectable by western blot in hypoxic $ZIP12^{-/-}$ rats but increased in hypoxic WT (WKY) rats ($n = 3$ per group). **e**, ZIP12 expression by immunohistochemistry of WT and $ZIP12^{-/-}$ rat lungs before and after hypoxia (2 weeks). **f**, ZIP12 expression in lungs from a chronic iron-deficient rat, monocrotaline (MCT) rat and a patient with IPAH. **g**, Double immunofluorescence demonstrates co-localization of ZIP12 and smooth muscle actin in the remodelled vessels from the patient with IPAH.

of tissue hypoxia²² (Extended Data Fig. 7). Interestingly, the F344 rat strain has previously been reported to exhibit some resistance to monocrotaline-induced pulmonary hypertension²³; this was recapitulated in the ZIP12^{-/-} rat (Extended Data Fig. 8). These data signal a fundamental role for ZIP12 in the regulation of pulmonary vascular homeostasis in hypoxic stress, relevant to the pathogenesis of pulmonary hypertension beyond that associated with life in a low-oxygen atmosphere. The current treatments for pulmonary hypertension centre on the pharmacological manipulation of signalling mechanisms used by vasoactive factors and have limited therapeutic benefit. Our observations open a new avenue of research into the therapeutic potential of ZIP12 inhibition and suppressed excursions of intracellular free zinc as a novel strategy for preventing or treating pulmonary hypertension.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 December 2014; accepted 3 June 2015.

Published online 10 August 2015.

- Wilkins, M. R. *et al.* Pathophysiology and treatment of high-altitude pulmonary vascular disease. *Circulation* **131**, 582–590 (2015).
- Zhao, L. *et al.* Right ventricular hypertrophy secondary to pulmonary hypertension is linked to rat chromosome 17: evaluation of cardiac ryanodine Ryr2 receptor as a candidate. *Circulation* **103**, 442–447 (2001).
- Rhodes, J. Comparative physiology of hypoxic pulmonary hypertension: historical clues from brisket disease. *J. Appl. Physiol.* **98**, 1092–1100 (2005).
- Atanur, S. S. *et al.* Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* **154**, 691–703 (2013).
- Chowanadisai, W. *et al.* Neurulation and neurite extension require the zinc transporter ZIP12 (slc39a12). *Proc. Natl Acad. Sci. USA* **110**, 9903–9908 (2013).
- Liuzzi, J. P. & Cousins, R. J. Mammalian zinc transporters. *Annu. Rev. Nutr.* **24**, 151–172 (2004).
- Schermuly, R. T. *et al.* Mechanisms of disease: pulmonary arterial hypertension. *Nature Rev. Cardiol.* **8**, 443–455 (2011).
- Chowanadisai, W. Comparative genomic analysis of slc39a12/ZIP12: insight into a zinc transporter required for vertebrate nervous system development. *PLoS ONE* **9**, e111535 (2014).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Vinkenborg, J. L. *et al.* Genetically encoded FRET sensors to monitor intracellular Zn²⁺ homeostasis. *Nature Methods* **6**, 737–740 (2009).
- Cui, X. *et al.* Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nature Biotechnol.* **29**, 64–67 (2011).
- Bernal, P. J. *et al.* A role for zinc in regulating hypoxia-induced contractile events in pulmonary endothelium. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **300**, L874–L886 (2011).
- Zhang, Y. *et al.* ZIP4 regulates pancreatic cancer cell growth by activating IL-6/STAT3 pathway through zinc finger transcription factor CREB. *Clin. Cancer Res.* **16**, 1423–1430 (2010).
- Grattan, B. J. & Freake, H. C. Zinc and cancer: implications for LIV-1 in breast cancer. *Nutrients* **4**, 648–675 (2012).
- Chen, Q. G. *et al.* The role of zinc transporter ZIP4 in prostate carcinoma. *Urol. Oncol.* **30**, 906–911 (2012).
- Zhao, L. *et al.* Sildenafil inhibits hypoxia-induced pulmonary hypertension. *Circulation* **104**, 424–428 (2001).
- Zhao, L. *et al.* Histone deacetylation inhibition in pulmonary hypertension: therapeutic potential of valproic acid and suberoylanilide hydroxamic acid. *Circulation* **126**, 455–467 (2012).
- Francis, S. H. *et al.* Zinc interactions and conserved motifs of the cGMP-binding cGMP-specific phosphodiesterase suggest that it is a zinc hydrolase. *J. Biol. Chem.* **269**, 22477–22480 (1994).
- Wharton, J. *et al.* Antiproliferative effects of phosphodiesterase type 5 inhibition in human pulmonary artery cells. *Am. J. Respir. Crit. Care Med.* **172**, 105–113 (2005).
- Cotroneo, E. *et al.* Iron homeostasis and pulmonary hypertension: iron deficiency leads to pulmonary vascular remodeling in the rat. *Circ. Res.* **116**, 1680–1690 (2015).
- Tuder, R. M. *et al.* Expression of angiogenesis-related molecules in plexiform lesions in severe pulmonary hypertension: evidence for a process of disordered angiogenesis. *J. Pathol.* **195**, 367–374 (2001).
- Beasley, N. J. *et al.* Carbonic anhydrase IX, an endogenous hypoxia marker, expression in head and neck squamous cell carcinoma and its relationship to hypoxia, necrosis, and microvessel density. *Cancer Res.* **61**, 5262–5267 (2001).
- Pan, L. C. *et al.* Strain differences in the response of Fischer-344 and Sprague-Dawley rats to monocrotaline induced pulmonary vascular-disease. *Toxicology* **79**, 21–35 (1993).

Acknowledgements This research was supported by successive grants from British Heart Foundation to M.R.W. and L.Z. (PG/95170, PG/98018, PG/2000137, PG/04/035/16912, PG/12/61/29818, PG/10/59/28478 and RG/10/16/28575). G.A.R. was supported by a Wellcome Trust Senior Investigator Award (WT098424AIA), MRC Programme Grant (MR/J0003042/1) and a Royal Society Research Merit Award. T.A. acknowledges support from European Research Council Advanced Grant ERC-2010-AdG, number 268880. We thank A. I. Garcia-Diaz for advice on genotyping techniques and R. Edwards for advice on antibody production. We thank C. Haley for discussions on the rat genetic map.

Author Contributions L.Z. and M.R.W. were principal investigators on grants from the British Heart Foundation, developed concepts and supervised the project. L.Z., M.R.W. and E.O. designed and implemented the experiments. T.A. gave conceptual advice on the congenic program and whole-genome sequencing. L.Z., M.R.W., T.A., K.M., E.O. and O.D.D. conducted the congenic breeding program. S.S.A. analysed whole-genome sequence data and performed the Polyphen analysis. L.Z., E.O. and O.D.D., with the support of B.M. and Z.W., generated the ZIP12 transgenic rat. L.Z., E.C. and L.W. performed immunohistochemistry and immunofluorescence. E.O. conducted the *in vitro* cell culture experiment. C.A. and E.O. performed the angiogenesis assay. G.R. supervised and E.O. and P.L.C. conducted the intracellular labile zinc measurement experiments. C.-N.C. and E.O. performed ChIP-PCR. J.P.-C. and E.O. cloned the HRE construct and performed luciferase reporter assays. M.G.F., K.R.S. and A.A. provided cattle and human lung sections. E.O. performed statistical analysis. L.Z., M.R.W. and E.O. interpreted the data and wrote the manuscript. T.A., G.R., J.F., S.S.A. and K.D. edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.Z. (l.zhao@imperial.ac.uk).

METHODS

Animals. Inbred Wistar-Kyoto (WKY, Charles River) and Fischer 344 rats (F344, Harlan) were used as original strains. Animals were maintained at a constant temperature (20–22 °C) with a 12-h on/12-h off light cycle, with food and water *ad libitum*. All experiments were conducted under the project licence PPL70/7425 in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986. To ascertain the pulmonary artery pressure phenotype, male rats (WKY, F344, congenic and sub-congenic strains, ZIP12 transgenic rats) aged 10–12 weeks were studied in batches, with the parental WKY strain included as an internal control in each batch studied. Sample sizes were chosen on the basis of experience of pulmonary artery pressure variation in the parental strains. A sample size of at least $n = 5$ per group was predicted to detect a difference in mean pulmonary arterial pressure³ of 5 mm Hg (SD = 3) with 95% power with 95% confidence. Additional animals were studied to obtain sufficient tissue for supportive analyses.

Generation of congenic and sub-congenic strains. To investigate the involvement of the chromosome 17 QTL in the pulmonary hypertension (PH) phenotype, we introgressed the F344 chromosome QTL segment into the WKY genetic background by repeated backcrossing². We produced a congenic rat strain, R47A (WKY.F344-D17Got91/D17Rat51), which contains 15 Mbp from the F344 donor region that maps to the distal end of the QTL on the WKY background.

Subsequently, we generated three sub-congenic strains: Sub-A (WKY.F344-D17Got91/D17Rat47), Sub-B (WKY.F344-D17Rat47/D17Rat51) and Sub-C (WKY.F344-D17Rat131/D17Rat51). These three recombination events divided the R47A congenic interval into three smaller and overlapping sub-congenic intervals (Extended Data Fig. 1).

Microsatellite genotyping of congenic rats. Congenic and sub-congenic rats were genotyped using simple sequence length polymorphism markers (Extended Data Table 2). To reduce the unknown regions between the markers, rats were also genotyped using primers specifically designed to amplify known regions containing insertions or deletions in one of the two parental strains (Extended Data Table 2). Genomic DNA was isolated from rat ear clippings using hot sodium hydroxide and Tris (HotSHOT) extraction²⁴. Forward primers were fluorescently labelled with 6-FAM. PCR products together with the fluorescent size marker (ROX 400HD, Applied Biosystems) were diluted in formamide and run on a 3730xl DNA Analyzer (Applied Biosystems). Results were analysed using GeneMapper version 3.7 software (Applied Biosystems).

Illumina genome sequencing library preparations. Five micrograms of male WKY/Ncr1 (two animals) and F344/Ncr1 (one animal) rats were used to construct paired-end whole-genome libraries with 300–550 bp insert size. Genomic DNA was prepared by standard phenol chloroform extraction followed by treatment with DNase-free RNase. DNA quality was assessed by spectrophotometry (260/280 and 260/230) and gel electrophoresis before library construction. Genomic DNA was sheared for 90 s (Covaris S2, KBioscience), using 10% duty cycle, 5 intensity and 200 cycles per burst. The shearing efficiency was assessed by Qubit 2.0 fluorometer measurements (Life Technologies) and gel electrophoresis. The library was prepared as recommended (Illumina Genomic DNA sample prep kit protocol) with nine cycles of PCR amplification (Illumina). Constructed libraries were assessed with an Agilent 2100 bioanalyser using an HS DNA assay (Agilent Technologies) and quantified using a KAPA Illumina SYBR Universal Lib QPCR kit (Anachem). The resulting libraries were sequenced on an Illumina HiSeq2000 following the manufacturer's instructions. Polymorphisms were confirmed by capillary sequencing.

Generation and genotyping of WKY.*Slc39a12*^{+/-} and *Slc39a12*^{-/-} rats. CompoZr Custom Zinc Finger Nucleases targeting the rat *Slc39a12* gene were designed and purchased from Sigma-Aldrich (Extended Data Fig. 3). Pronuclei from fertilized WKY oocytes were microinjected with ZFN mRNA (2 ng μ l⁻¹). Three out of eleven pups were positive for *Slc39a12* mutations, as revealed by Cel-I surveyor assay and gene sequencing¹¹. One pup (mutant 77) hosted a stop codon 15 amino acids from the ZFN-binding site, resulting in a truncated protein of 490 amino acids (54 kDa), 198 amino acids smaller than the WT protein, and introduced a sequence coding for 5'-ATTAAAT-3', a binding site for the SmaI restriction enzyme. Mutant 77 was selected as a founder to mate with a WKY female. Pups were genotyped by amplifying DNA and digesting with SmaI. The primers used to amplify the region of interest were forward 5'-GCAATG-GTTTCCACAGTGA-3' and reverse 5'-GCGCACTGAGGCTTTAAGAA-3'.

Pulmonary hypertension phenotyping. Pulmonary hypertension was induced by placing animals in a normobaric hypoxic chamber (FIO₂ = 10%) for 2 weeks or by subcutaneous injection of monocrotaline (60 mg kg⁻¹; Sigma-Aldrich). The experiments were not randomized. All studies were performed using the same equipment and all haemodynamic measurements were made by the same two operators. The operators were blinded to the genetic status of animals, which was confirmed after phenotyping. All histological assessments were made by two observers blinded to genetic status of the animals. At the end of each experimental

period, animals were weighed and anaesthetized (Hypnorm 1 ml kg⁻¹ intramuscularly; Mydazolam 0.8 ml kg⁻¹ intraperitoneally). Pulmonary arterial pressure was measured with a pre-curved catheter inserted through the right jugular vein. Systemic blood pressure was assessed via carotid artery cannulation. Cardiac output was measured by thermodilution. PVR was calculated using the standard equation, PVR = mean pulmonary artery pressure/cardiac output. All data were recorded with a PowerLab Data Acquisition system (AD Instruments) and analysed using LabChart 7 software.

The animals were then killed and the heart dissected, and individual chamber weights recorded. The ratio of right ventricle to left ventricle plus the septum mass was calculated as the right ventricle hypertrophy index. Some collected tissues were snap frozen in liquid nitrogen and stored at -80 °C for further biochemical measurements. The left lung was fixed by inflation with 10% formalin in phosphate-buffered saline, embedded in paraffin and sectioned for histology. Transverse rat lung sections were processed for elastic van Gieson staining. Peripheral vessels less than 100 μ m in diameter were counted at $\times 40$ magnification objective, and pulmonary vascular remodelling was expressed as the proportion of vessels with double elastic lamina (>75% of the circumference as fully muscularized, 25–75% as partly muscularized) to total vessels counted. Counting was performed twice by observers blinded to treatment.

Ex vivo angiogenesis assay of pulmonary arteriole. Angiogenesis assay of arterial rings was performed as previously described²⁵. Pulmonary arterioles (first and second order) were dissected from rat lungs viewed under the microscope. One millimetre sections were placed in matrigel (50 μ l per well) in a 96-well plate, allowed to gel for 30 min at room temperature (20 °C), then incubated for up to 6 days with endothelial cell culture medium MV2 with 5% fetal calf serum (PromoCell). On days 3, 4, 5 and 6, the length of the longest sprouts was measured under the microscope ($\times 4$ objective). On day 6, fluorescent images of the arteriole ring were taken by staining the tissue with calcein (Invitrogen) for 15 min at 37 °C.

Anti-ZIP12 antibody production. An antibody raised against the last five amino acids at the C terminus of both the human and the rat ZIP12 protein was produced in rabbits following previous methodology²⁶. Rabbits were immunized with synthetic peptides conjugated to keyhole limpet haemocyanin ([CYS(KLH)QNIKI]. Peptide sequence was confirmed to be ZIP12 specific using RStudio. Immunized rabbit serum containing anti-ZIP12 antibody specificity was confirmed by immunoblotting with rat lung lysates or human pulmonary smooth muscle cells. A single band at about 70 kDa was visible in the immunoblots.

Lung immunohistochemistry and immunofluorescence. Human IPAH and control lung samples were obtained from the Imperial College pulmonary hypertension biorepository (ethics reference numbers 01-210 and 2001/6003) where samples are deposited following informed patient consent. Anonymized Kyrgyz high-altitude lung samples were obtained from post-mortem lung following approval of the local ethics committee (reference 02-23/880).

Lung sections were immunostained with rabbit anti-ZIP12 (1:1,000), Ki67 (1:50; Thermo Scientific) and rabbit anti-CAIX (1:100) antibodies. For immunohistochemistry, horseradish-peroxidase-conjugated secondary anti-rabbit antibody (1:200) was used. Double immunofluorescence with anti- α SMA (1:100) was performed using fluorescence secondary antibodies, anti-mouse Alexa 488 and anti-rabbit Alexa 568 (1:2,000, Invitrogen). Images (green for ZIP12 and red for α SMA) were detected under Leica confocal microscope (TCS SP2 AOBs).

HPASMC culture. HPASMCs from PromoCell and Lonza were grown in human smooth muscle cell growth medium 2 (PromoCell). All commercial cell lines were tested for mycoplasma contamination. The cells were cultured under normal oxygen tension (20% O₂, 5% CO₂) or exposed to hypoxia (2% O₂, 5% CO₂, 92% N₂) for 48–72 h. A bromodeoxyuridine (BrdU) cell proliferation assay (Millipore) was used to assess cell proliferation following the manufacturer's conditions.

ZIP12 siRNA transfection. Cells were transfected overnight with 50 pmol siRNA against ZIP12 (s8397, Ambion), or negative siRNA (4390844, Ambion) as a control, using Lipofectamine RNAiMAX (Invitrogen Life Technologies) according to the manufacturer's conditions.

Quantification of actin fibre formation. Cells were cultured on plastic coverslips (Nunc), transfected with scrambled or ZIP12 siRNA and exposed to hypoxia as described previously. After exposure for 48 h, cells were fixed with 4% formaldehyde solution in phosphate buffered saline for 10 min at room temperature. Cells were then incubated with Alexa-568-conjugated phalloidin (1/200; Invitrogen) for F-actin detection by confocal microscopy. Sequential *xyz*-sections (approximately 12 sections of 1 μ m² per view) were obtained and three-dimensional images were reconstructed. Quantification of actin stress fibres was determined by volume rendering in Image-J. Actin volume per cell was expressed as the fold increase from normoxic control (value set at 1).

Quantification of zinc concentration by FRET measurement. Cells on coverslips were washed twice in Krebs-HEPES-bicarbonate (KHB) buffer (140 mM NaCl, 3.6 mM KCl, 0.5 mM NaH₂PO₄, 0.2 mM MgSO₄, 1.5 mM CaCl₂, 10 mM

HEPES, 25 mM NaHCO₃), which was warmed, bubbled with 95% O₂/5% CO₂, set to pH 7.4 and contained 3 mM glucose. Imaging of zinc using eCALWY sensors was performed as optimized before^{10,27}. Briefly, cells were maintained at 37 °C throughout with a heating stage (MC60, LINKAM, Scientific Instruments), and KHB buffer was perfused (1.5–2 ml min⁻¹) with additions as stated in the figures. Images were captured at 433 nm monochromatic excitation wavelength (Polychrome IV, Till Photonics) using an Olympus IX-70 wide-field microscope with a ×40/1.35 numerical aperture oil immersion objective and a Zyla sCMOS camera (Andor Technology) controlled by Micromanager software²⁸. Acquisition rate was 20 images per minute. Emitted light was split and filtered by a Dual-View beam splitter (Photometrics) equipped with a 505dxcn dichroic mirror and two emission filters (Chroma Technology D470/24 for cerulean and D535/30 for citrine).

Image analysis used ImageJ software²⁹ with a homemade macro, and the fluorescence emission ratios were derived after subtracting background. Steady-state fluorescence intensity ratio of citrine/cerulean (*R*) was measured, then maximum and minimum ratios were determined to calculate free Zn²⁺ concentrations using the following formula: $[Zn^{2+}] = K_d \times (R_{max} - R)/(R - R_{min})$. The maximum ratio (*R*_{max}) was obtained upon intracellular zinc chelation with 50 μM TPEN and the minimum ratio (*R*_{min}) was obtained upon zinc saturation with 100 μM ZnCl₂ in the presence of the Zn²⁺ ionophore pyrithione (5 μM) (ref. 10).

HIF-motif analysis and cloning. HOMER⁹ was used to scan for HIF-1α and HIF-2α recognition motifs in the region 2 kb upstream and 1.5 kb downstream of the ZIP12 transcription start site. Results with a HOMER score <6.5 were discarded. A 5' region of ZIP12 gene containing these motif (HRE) (human (hg19) chromosome 10: positions 18,240,587–18,242,100) was cloned into the multicloning site of pGL4.10, which encodes the luciferase reporter gene *luc2*, by Gibson Assembly (NEB, E2611S). Three nucleotide substitutions in the core of the predicted HIF1/2α-binding site motif were created by site-mutagenic PCR to produce a disabling mutant (Fig. 2d).

Transfection and luciferase assay. HPASMCs were seeded in 24-well plates at 70–80% confluence. Cells were transfected with 300 ng of each plasmid together with 2 ng of Renilla plasmid using Lipofectamine 2000 (Life Technologies), exposed to hypoxia and lysed according to the manufacturer's conditions. Luciferase activity was measured using Dual-Luciferase Reporter Assay Chemistry (Promega) as previously described³⁰. Experiments were repeated in two cell lines, *n* = 6 per line.

ChIP and PCR. Specific protein–DNA interactions were examined by ChIP followed by qPCR (Chromatin Immunoprecipitation Assay Kit, Millipore). Protein–DNA crosslinks were achieved by fixation with 1% formaldehyde for 10 min at room temperature. DNA–protein complexes from 2 × 10⁶ cells were sheared to lengths between 200 and 500 base pairs by sonicator (Bioruptor). The pre-cleared fragments were incubated with 10 μg of HIF-1α or HIF-2α specific antibody (Novus Biologicals), or without antibody (as a negative control) overnight, followed by immunoprecipitation by Protein A Agarose/Salmon Sperm DNA (50% Slurry). The crosslinks were reversed by heating at 65 °C overnight, followed by Proteinase K digestion at 45 °C for 2 h. DNA was then recovered with QIAquick PCR purification kit (Qiagen) for qPCR to prove affinity against ZIP12 promoter region (Fig. 2d). Experiments were conducted in two separate cell lines (*n* = 3 each) and gave the same result.

Quantitative PCR was performed as previously described in the methodology, using 1 μl of DNA sample, and using the forward primer 5'-TTTCCCAACCTGGGTCCTAT-3 and the reverse primer 5'-AGCAGCCAAAAGCTTGCTA-3. C terminus values were normalized compared with the values detected in the starting non-immunoprecipitated DNA sample (input). Protein–DNA affinity

was confirmed when normalized C terminus values were above the basal levels measured in the negative control.

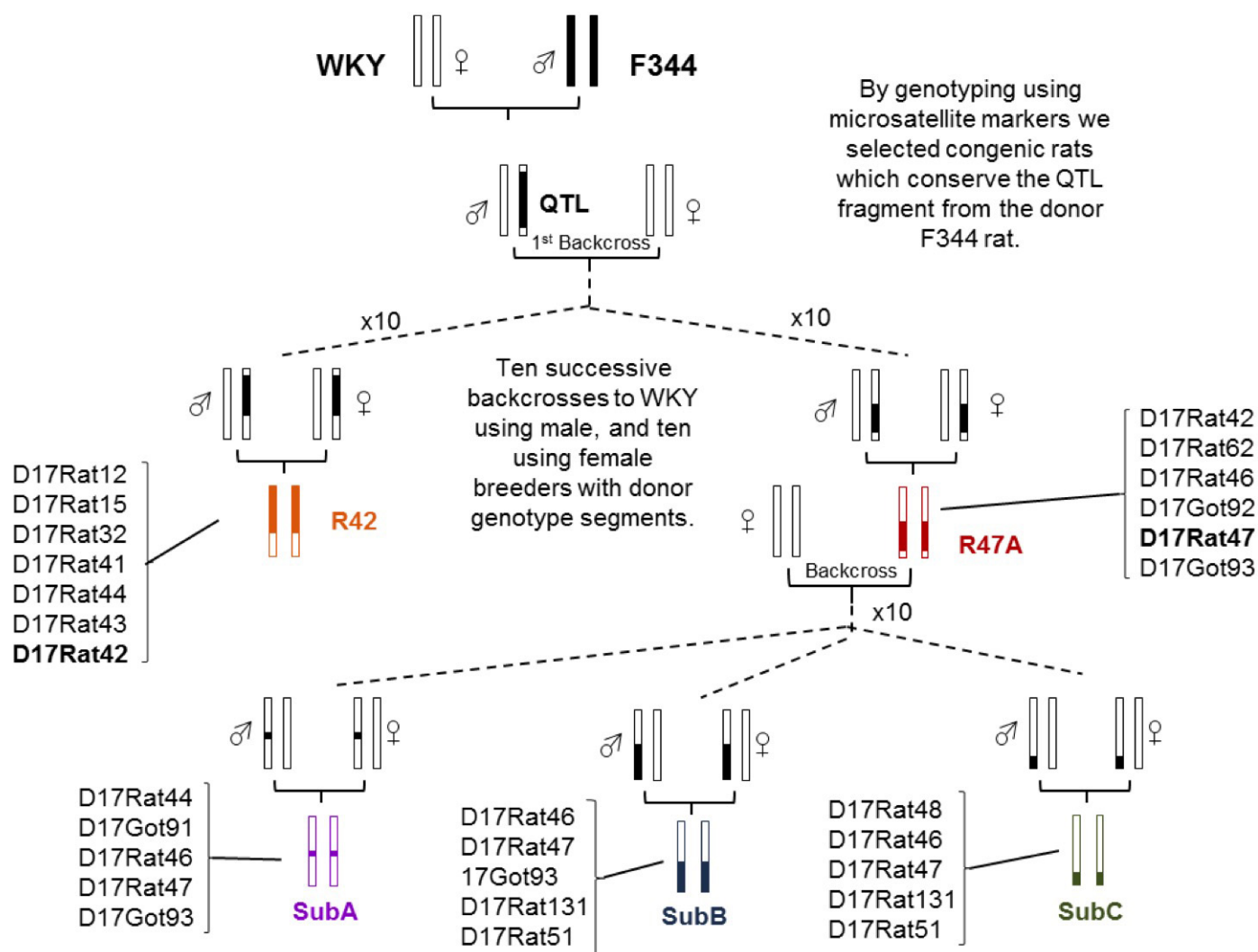
Quantitative PCR with reverse transcription. RNA was extracted from lungs using RNeasy Mini Kit (Qiagen). PCR was performed with an ABI 7500 Sequence Detection System (Applied Biosystems). Quantitative PCR was performed using a two-step protocol starting with complementary DNA (cDNA) synthesis using the ImProm-II Reverse Transcription System (Promega), followed by PCR using the Power SYBR Green PCR Master Mix (Applied Biosystems). A total of 100 ng of cDNA per sample was used. All samples were amplified using biological triplicates with two technical replicates per sample. The 7500 Sequence Detection System software (Applied Biosystems) was used to obtain C_T values. Results were analysed using the comparative C_T method³¹. Samples were normalized to a reference gene, *Ubc* (for rat samples) or *Cyclophilin* (for human cell samples), to account for cDNA loading differences.

Western blot. Frozen rat tissues (lungs) and cell pellets were homogenized in RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM sodium chloride, 1.0% Igepal, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate) (Sigma) supplemented with protease inhibitor (Roche). Western blotting used Mini-PROTEAN TGX Precast Gels (Bio-rad) following the manufacturer's suggestions. Blots were incubated for 1 h at room temperature with Anti-ZIP12 (1:10,000); Anti-HIF1α (1:1,000, Novus Biological) or Anti-HIF2α (1:1,000, Novus Biological). Proteins were detected by Clarity western ECL substrate (Bio-rad). Optical densities of individual bands were measured using ImageJ software and protein expressions were standardized with vinculin.

Statistical analysis. Data are presented as mean ± s.e.m. Data were tested for normality using a Kolmogorov–Smirnov test. All data were confirmed as normally distributed with similar variance between comparator groups. Data were analysed using one-way ANOVA followed by Bonferroni post-test adjustment for multiple comparisons or unpaired *t*-test. Graphpad Prism was used for all statistical analysis.

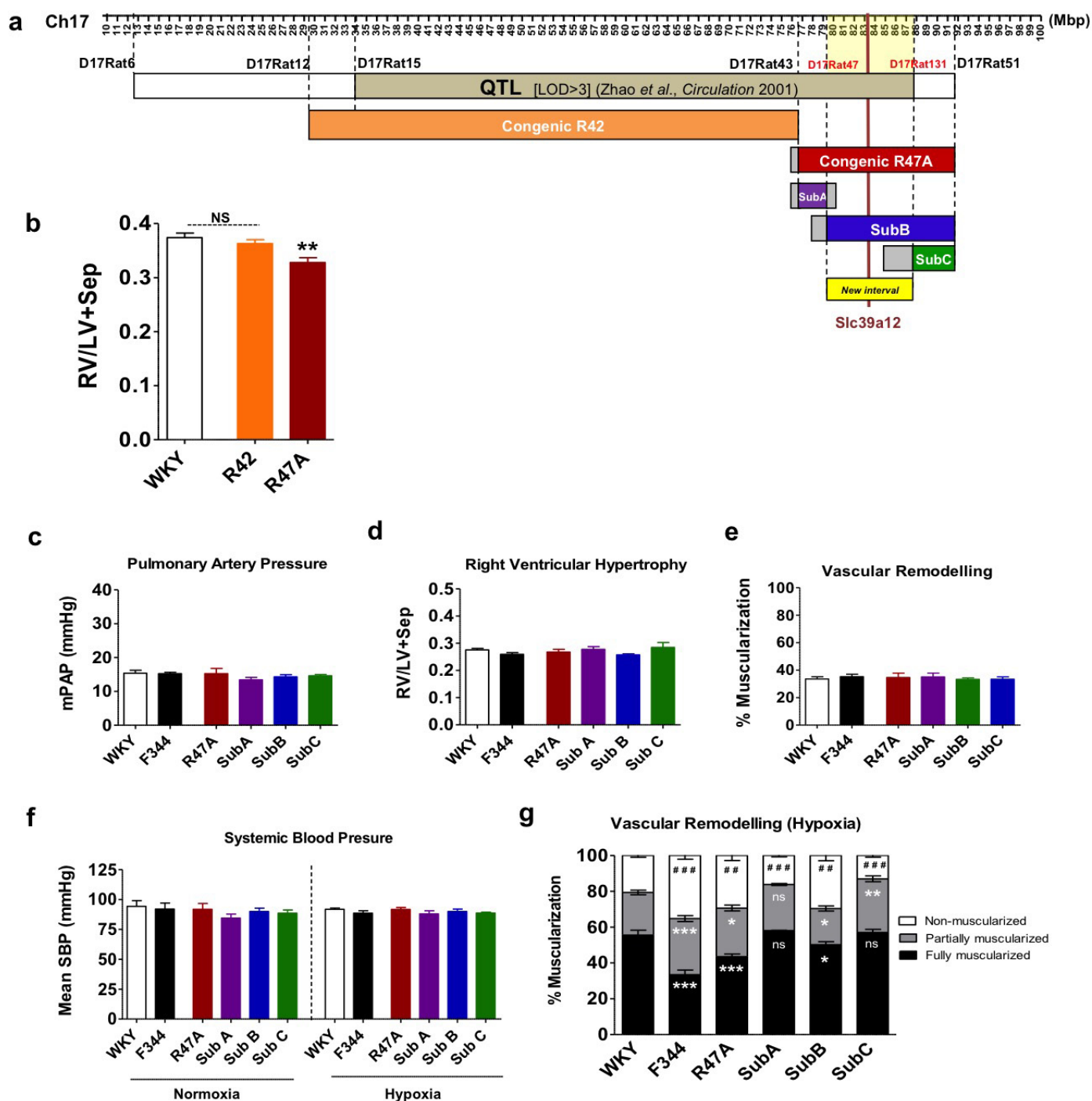
Other bioinformatics analyses. The Ensembl database³² was mined with the BioMart tool³³ to identify all transcribed elements in the confidence interval region. The search was limited to chromosome 17 between positions 85,072,475 and 93,347,758. PolyPhen analysis was used to predict the possible impact of described SNP on amino-acid substitution on the structure and function of a human protein³⁴.

24. Truett, G. E. *et al.* Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**, 54 (2000).
25. Aplin, A. C. & Nicosia, R. F. The rat aortic ring model of angiogenesis. *Methods Mol. Biol.* **1214**, 255–264 (2015).
26. Edwards, R. J. Targeting antipeptide antibodies toward cytochrome P450 enzymes. *Methods Mol. Biol.* **320**, 173–182 (2006).
27. Bellomo, E. A. *et al.* Glucose regulates free cytosolic Zn²⁺ concentration, Slc39 (ZIP), and metallothionein gene expression in primary pancreatic islet β-cells. *J. Biol. Chem.* **286**, 25778–25789 (2011).
28. Edelstein, A. *et al.* in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) Ch. 14, Unit 14.20 (Wiley, 2010).
29. Schneider, C. A. *et al.* NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
30. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nature Genet.* **46**, 136–143 (2014).
31. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3**, 1101–1108 (2008).
32. Flicek, P. *et al.* Ensembl's 10th year. *Nucleic Acids Res.* **38**, D557–D562 (2010).
33. Haider, S. *et al.* BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.* **37**, W23–W27 (2009).
34. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).



Extended Data Figure 1 | Generation of congenic and sub-congenic strains. Congenic rat lines were produced by introgression of the F344 chromosome 17 QTL segment onto the WKY genetic background by repeated backcrossing. Congenic rat strain R47A (WKY.F344-D17Got91/D17Rat51) contains 15 Mbp from the F344 donor region that maps to the distal end of the QTL on a WKY background. Three sub-congenic strains, SubA (WKY.F344-D17Got91/

D17Rat47), SubB (WKY.F344-D17Rat47/D17Rat51) and SubC (WKY.F344-D17Rat131/D17Rat51), were produced containing separate fragments of the R47A donor region by backcrossing (R47A × WKY) F₁ with WKY parental rats. Three recombination events within the R47A congenic interval divided the congenic interval into three smaller and overlapping sub-congenic intervals (Fig. 1 and main text).

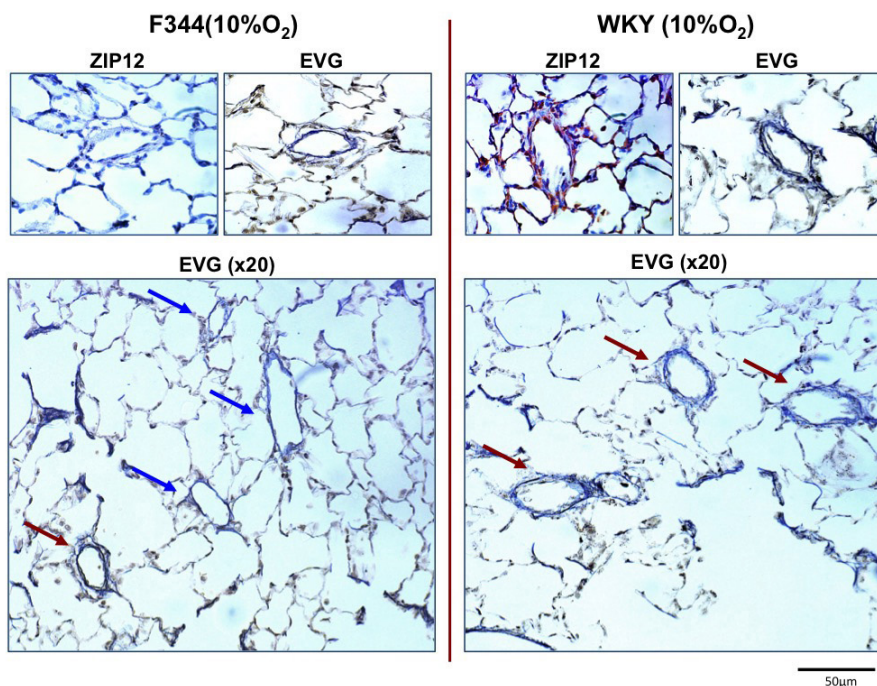


Extended Data Figure 2 | Dissection of QTL and cardiovascular phenotype of rat strains. **a**, An illustrative genetic map showing the relationship of the congenic strains (R42, R47A), subcongenic strains (SubA, SubB, SubC) and Slc39a12 to the original QTL (defined by a LOD score >3 (ref. 2)) on a physical map of chromosome 17 (using Rat Genome Assembly version 5.0). **b**, The hypoxia-resistant F344 phenotype tracks with the congenic R47A line. Rats were kept in 10% O₂ for 2 weeks and right ventricular hypertrophy was significantly attenuated in the congenic R47A strain (0.32 ± 0.03 , $n = 13$, $**P < 0.01$) compared with WKY rats (0.37 ± 0.03 , $n = 15$), whereas congenic R42 rats (0.36 ± 0.03 , $n = 17$) were similar (NS) to WKY rats. In normoxia, WKY, F344, R47A, SubA, SubB and SubC rats show no significant differences

in (c) mPAP, (d) right ventricular hypertrophy and (e) vascular muscularization ($n = 8$ each group); **f**, systemic blood pressure (SBP) is similar in all strains in both normoxia and hypoxic conditions. **g**, F344, R47A and SubB rats exhibit attenuated pulmonary vascular remodelling after 2 weeks exposure to a 10% O₂ atmosphere compared with WKY, SubA and SubC rats ($n = 6$ each group). Values are expressed as the mean \pm s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ compared with WKY (percentage of fully muscularized and partly muscularized vessels); $##P < 0.01$, $###P < 0.001$ compared with WKY (percentage of non-muscularized vessels) after one-way ANOVA followed by Bonferroni correction for multiple testing.

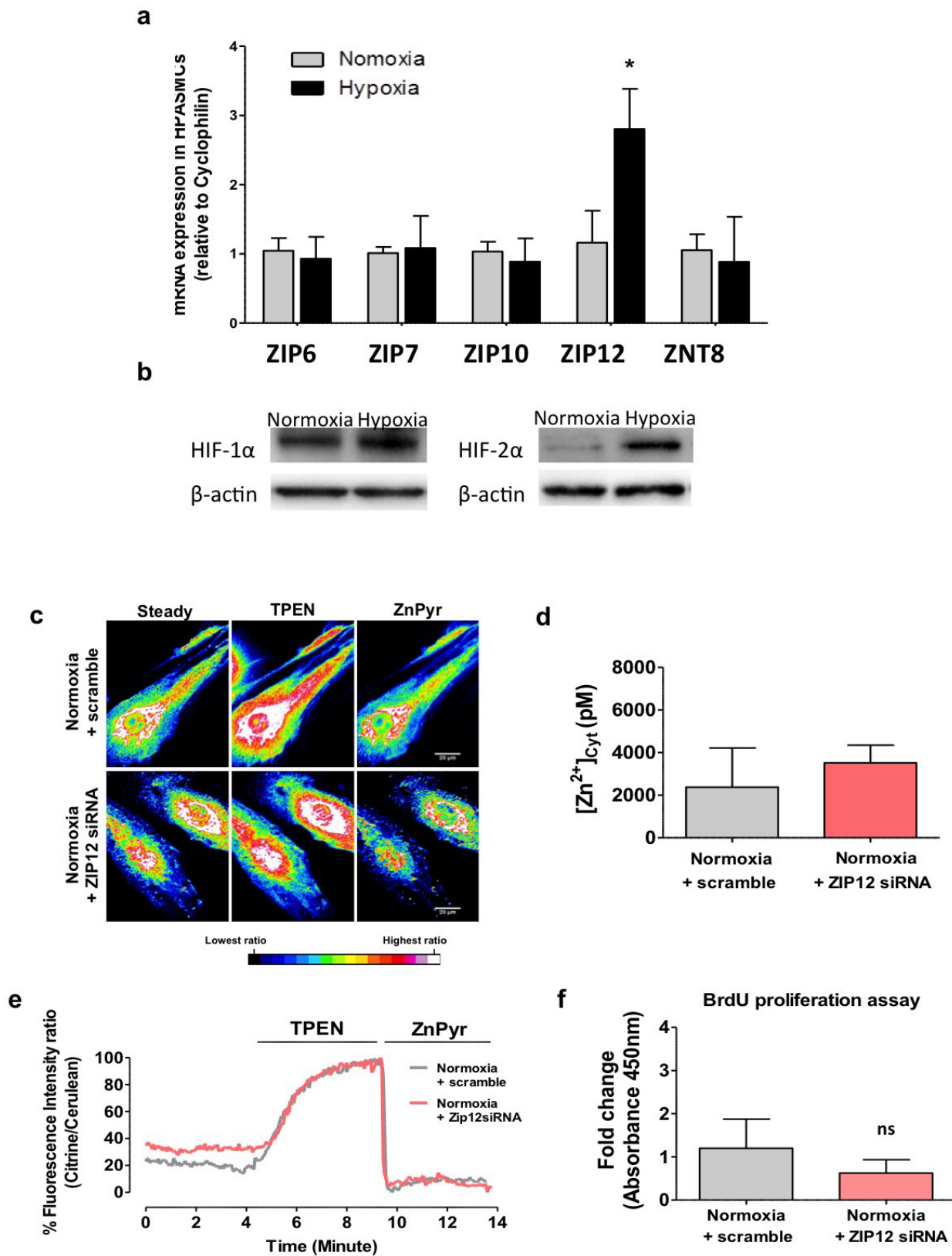
a

1	MCFWTELSVWVVLVSHSFSLASSTETSKALTQDNSRVGSHSLLEVLRLVLSAGDDRSLNHP	60
61	QSLIKILLERTGCPQRTDGTQEDCKLCLEPDALSLTAGGDLEDELREEVQRVSLLLLY	120
121	IIHQEEICSSKLNMSNREYEFYLSLLSLRQDEDSYFLSEKETDDILASTRKYFGTSTSP	180
181	CMETKILQRESGIQSSNGADEKMLPQLAATIIALSQGVCLGRKDSPPDDFTEYIFSFL	240
241	NRTNALHLSELELLNMLSTRRACTKINTLHEHQRKQNTAVHGLRDPKSAAMDVKVSGDH	300
301	SVSWDQACFSAQQLVEIFLQNHSSLSISKEDFKQLSPGIIQQLSCSCQVPRDQKAKPPP	360
361	TTLEKYGYSTVAVTLLTGLSMLGTALVLFHSCSEENYSLILQLFVGLAVGTLSGDALLHLI	420
421	PQVLGLHKQEAELGHFHESQSPIWKLLGLLGGIHGFFLIEKCFILLVSPNTKGLPLVNGH	480
481	AGHTHHLGLSPELNDQSGGKSIISTIQKGPEDSQTAELPKGNVPASNNRRTISLLAVM	540
541	VLVGDGLHNFADGLVIGTAFSSSLESGVTTTIAILC HEIPHEMGD FAVLLSSGLSIRTAI	600
	VLVEMACTILPMA.....	
601	LMNFLSALTAFIGLYIGLSVSADPRVQDWILTVTAGMFLYLSLVGMLPEMTHVQTRPMM	660
661	TFLNQVGLVGLWFSLLLLAVYEQNIKI.	688
	

b

Extended Data Figure 3 | Hypoxia-induced pulmonary vascular remodelling in parental strains. **a**, Upper panel sequence shows the WKY protein sequence (688aa); lower panel shows the truncated F344 protein sequence (553aa). Stars (*) mark the mutated amino acids compared with WKY protein. Dotted line indicates the C-terminal truncated region in F344.

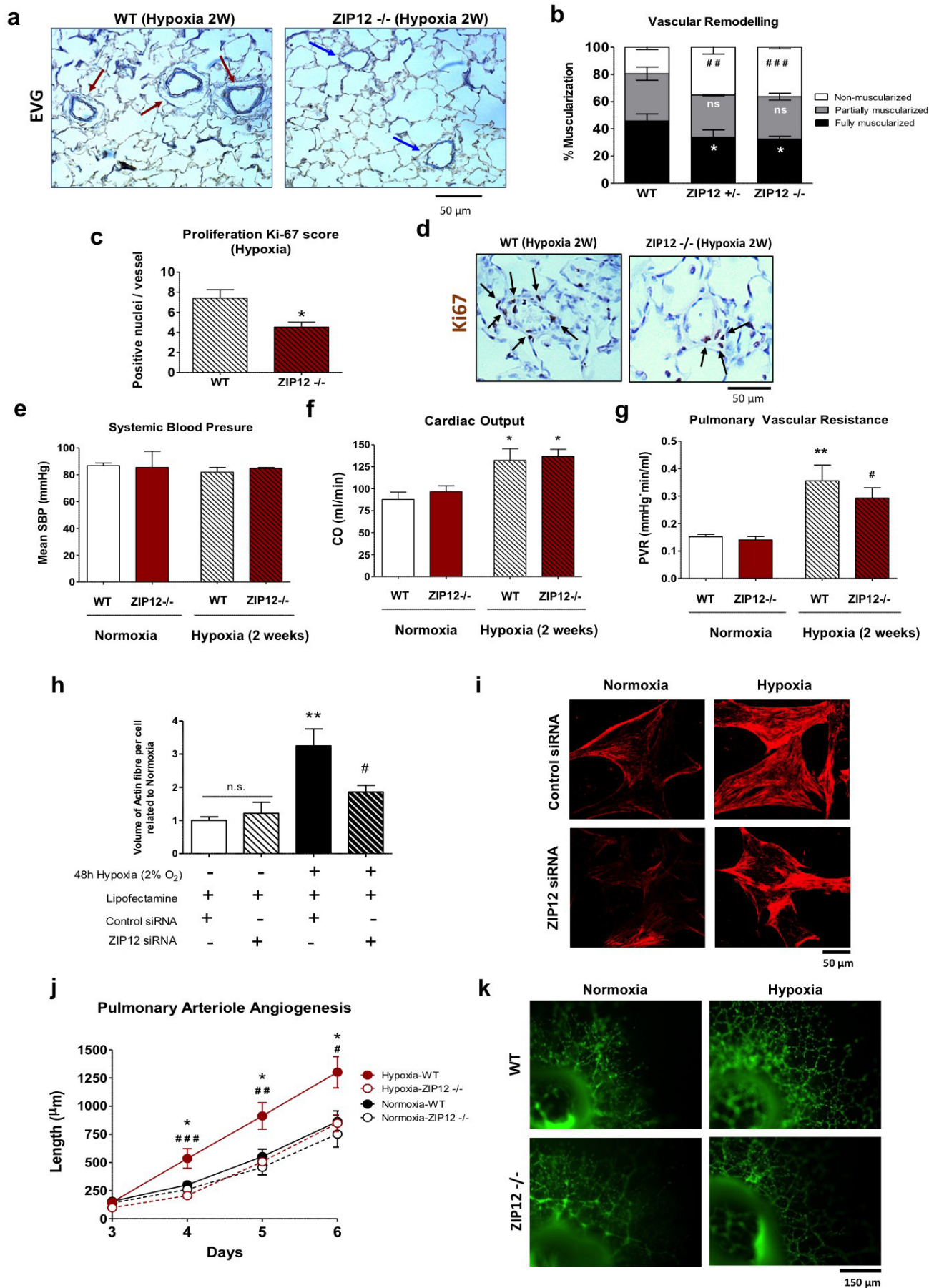
The grey square highlights the metalloprotease motif. **b**, Prominent ZIP12 immunostaining is seen in remodelled pulmonary arterioles in the chronically hypoxic WKY rat alongside vessels with a double elastic lamina (stained with elastic Van Gieson) but not F344 lungs exposed to hypoxia. Red arrow, vessel with double elastic lamina; blue arrow, vessel with single elastic lamina.



Extended Data Figure 4 | ZIP12 upregulation in response to hypoxia exposure and measurements of intracellular labile zinc concentration and proliferation of HPASMCs in normoxic conditions. **a**, Upregulation of ZIP12 in HPASMCs exposed to hypoxia, in contrast to other zinc transporters ($n = 6$). **b**, Representative western blots demonstrating increased HIF-1 α and HIF-2 α expression in HPASMCs after exposure to hypoxia for 24 h. **c**, Confocal

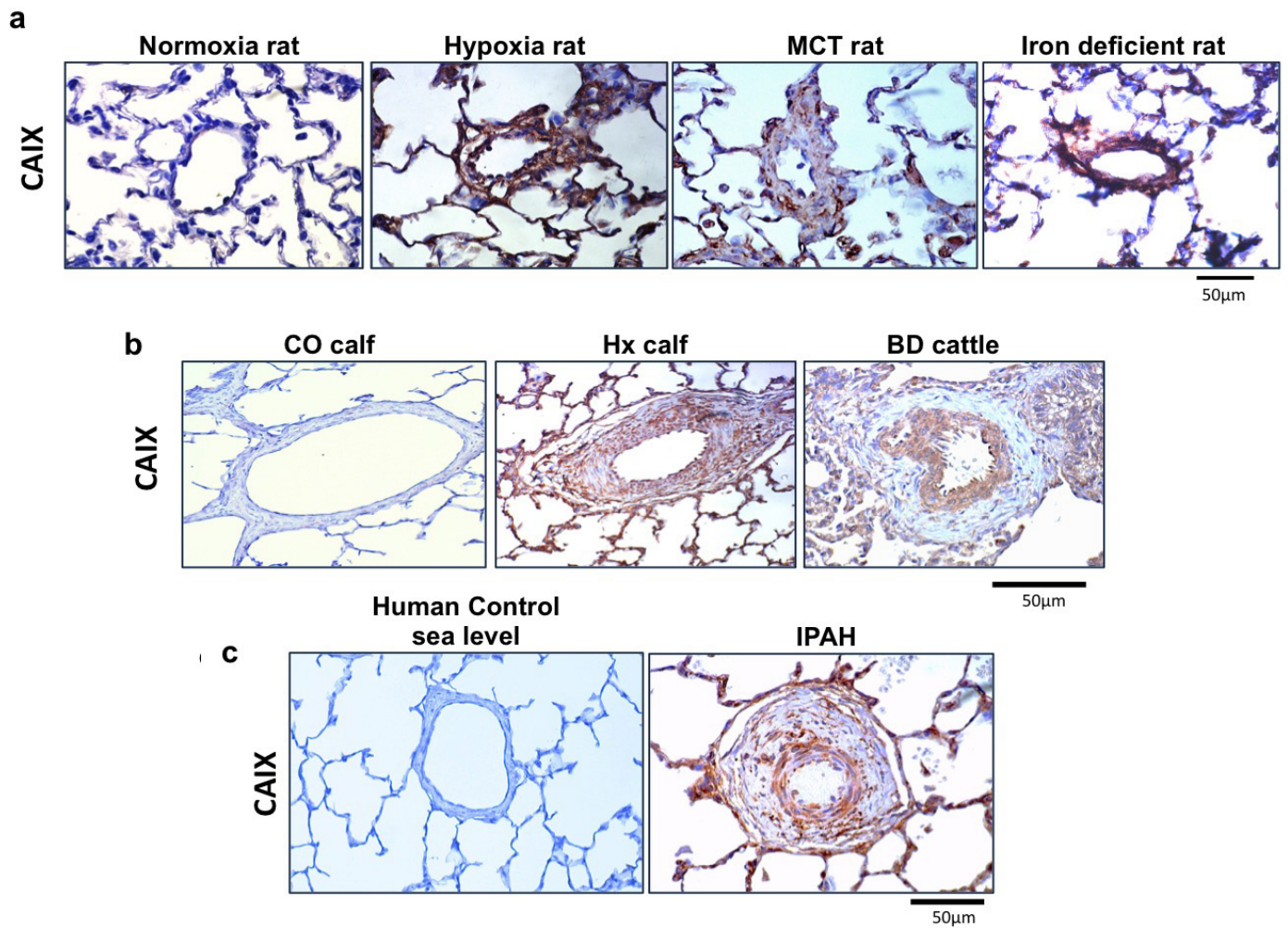
laser scanning images of HPASMCs transfected with eCALWY-4 probe. Intracellular free zinc was not affected by transfection with ZIP12 siRNA in normoxia. **d**, Representative traces showing the changes in fluorescence ratio using the eCALWY-4 probe. **e**, Quantification of intracellular zinc levels ($n = 10$). **f**, ZIP12 siRNA did not affect proliferation of HPASMCs in normoxic conditions ($n = 5$).

Extended Data Figure 5 | Design of specific Slc39a12 ZFN and confirmation of mutant line. **a**, CompoZr Custom Zinc Finger Nucleases (Sigma-Aldrich) for the rat Slc39a12 gene were designed to target exon 8. **b–d**, Cel-I surveyor assay and gene sequencing confirmed NHEJ-induced mutations in at least one pup (mutant 77). **e**, The 4-bp (AGTT) deletion followed by 2-bp insertion (TA) into mutant 77 caused a frame-shift in coding, introducing a stop codon leading to a truncated protein. Red star refers to stop codon. **c**, We subsequently genotyped next generation litters using SwaI (cutting point: 5'-ATTTAAAT-3'), showing 100% digestion for homozygous pups (−/−), 50% for heterozygous (+/−) and no DNA digestion for WT rats (+/+).



Extended Data Figure 6 | ZIP12 knockout attenuated hypoxia-induced pulmonary vascular remodelling. **a**, Representative lung sections from WT and ZIP12^{-/-} rats 2 weeks after hypoxia exposure. Elastic van Gieson staining showing double elastic lamina (red arrow) in WT but single elastic laminae (blue arrow) in ZIP12^{-/-} rats. **b**, Genetic disruption of ZIP12 in WKY rat attenuated pulmonary vascular remodelling after exposure for 2 weeks to a 10% O₂ atmosphere compared with WT rats ($n = 5$ each group). * $P < 0.01$ compared with WT (percentage of fully muscularized vessels); ## $P < 0.01$, ### $P < 0.001$ compared with WT (percentage of non-muscularized vessels) after one-way ANOVA followed by Bonferroni's multiple comparison test. **c**, Ki67 staining showing reduced proliferation in hypoxic ZIP12^{-/-} rat lungs compared with the WT strain. * $P < 0.01$ compared with WT. **d**, Representative sections from hypoxic WT and ZIP12^{-/-} rats lungs showing differences in staining with the proliferation marker, Ki67. **e–g**, Genetic disruption of ZIP12 in WKY rat did not influence **(e)** systemic blood pressure (SBP) or **(f)** cardiac

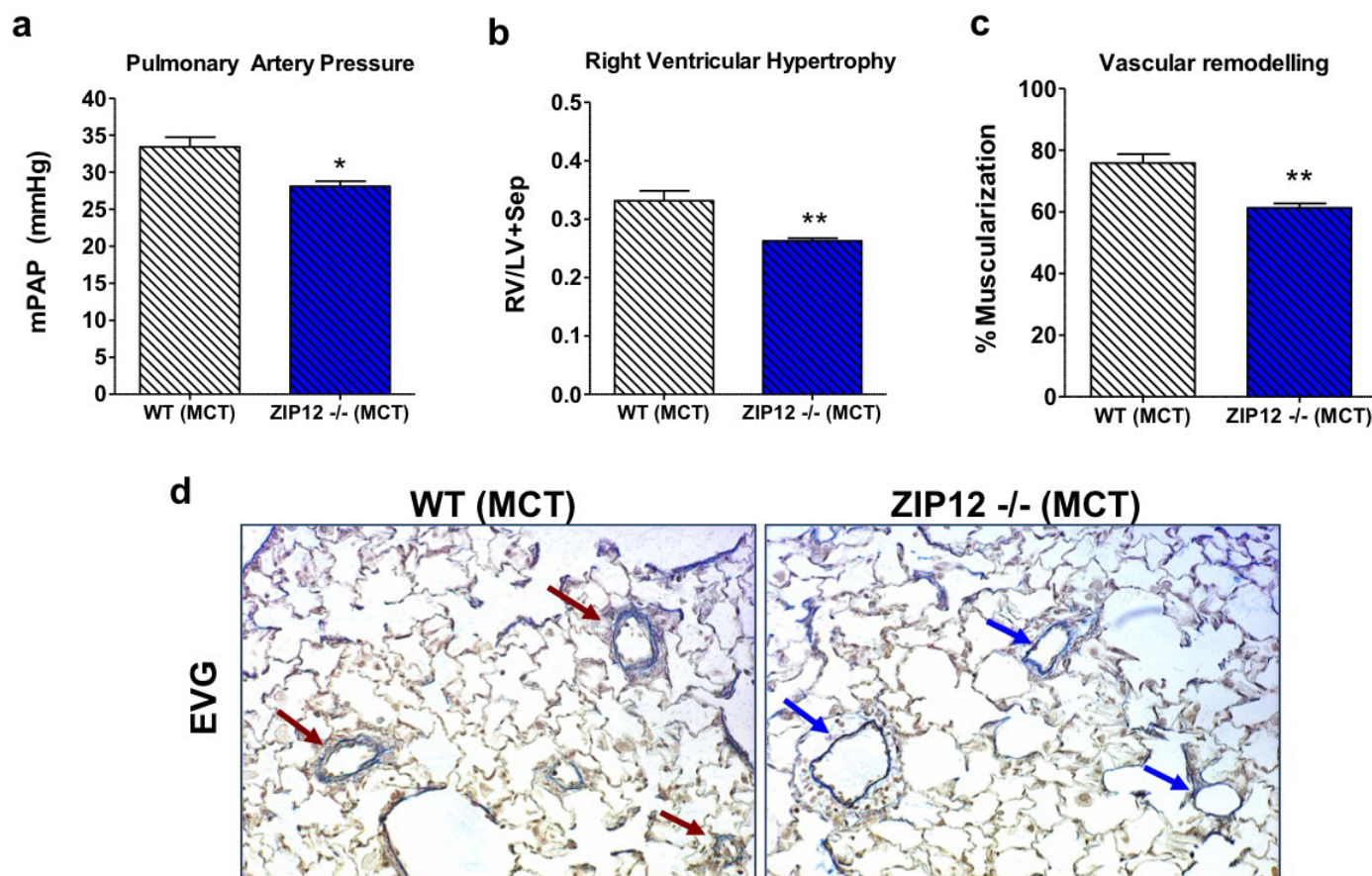
output (CO) but attenuated hypoxia-induced increases in **(g)** PVR ($n = 7$ each group). Values are expressed as the mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ compared with normoxic rats, # $P < 0.05$ compared with WT hypoxic rats after one-way ANOVA followed by Bonferroni correction for multiple testing. **h**, ZIP12-targeted siRNA inhibition attenuates stress fibre formation in HPASMCs in hypoxia ($n = 5$ each group). ** $P < 0.01$ compared with normoxia control group, # $P < 0.05$ compared with hypoxia control group. **i**, Representative pictures of actin stress fibre in HPASMCs. **j**, *Ex vivo* angiogenesis studies demonstrated that vascular outgrowth from ZIP12^{-/-} pulmonary vessels in response to hypoxia was attenuated compared with vessels from WT rats ($n = 12$ each group, 2 rings per rat, 6 ZIP12^{-/-} and 6 WT rats). * $P < 0.05$ compared with normoxia WT group; # $P < 0.05$, ## $P < 0.01$ and ### $P < 0.001$ compared with hypoxia ZIP12^{-/-} group. **k**, Representative pictures of pulmonary arteriole ring outgrowth at day 6.



Extended Data Figure 7 | Carbonic anhydrase (CAIX) expression.

a, Representative sections demonstrating increased CAIX expression in remodelled pulmonary arterioles in the lungs of rats exposed to alveolar hypoxia (2 weeks), monocrotaline (MCT, 3 weeks) or a chronic iron-deficient diet (4 weeks). **b, c**, No CAIX staining was detected in pulmonary arteries of

low-altitude (normoxia control, CO calf) calves and sea-level humans, but prominent CAIX immunostaining was observed in the remodelled pulmonary arteries of calves with severe pulmonary hypertension (Hx calf), in cattle with naturally occurring pulmonary hypertension ('Brisket disease', BD) as well as patients with IPAH.



Extended Data Figure 8 | Genetic disruption of ZIP12 in WKY rat attenuated monocrotaline-induced pulmonary hypertension. **a**, mPAP, **(b)** right ventricular hypertrophy and **(c)** pulmonary arteriole muscularization ($n = 5$ each group). Values are expressed as the mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ compared with WT monocrotaline group after unpaired Student's

t -test. **d**, Representative lung sections from WT and ZIP12^{-/-} rats 3 weeks after monocrotaline injection. Elastic van Gieson staining showing double elastic lamina (red arrow) in WT but single elastic laminae (blue arrow) in ZIP12^{-/-} rats.

Extended Data Table 1 | Frameshift and non-synonymous coding mutations in the refined congenic interval of F344 and the other hypoxia-susceptible strains, WKY, spontaneously hypertensive (SHR) and fawn-hooded (FHH) rat strains

Gene Name	ENSEMBL gene ID	ENSEMBL transcript ID	SNP/Indel ID	Chromosome	Position		Reference Allele	Alternate Allele	Strain							Consequences	cDNA position	CDS position	Protein position	Amino acid change	Codon Change	Polyphen Prediction	
					Start	End			F344/NCrl	FHH/funkheei	SHR/Olahece	SHR/Nind	SHRSP/ela	WKY/NCrl	WKY/ela								WKY/Nind
Cdrlf	ENSRNOG000000026493	ENSRNOT000000036299	17_85839161_G/A	17	85,839,161	-	G	A	2	0	0	0	0	6	0	0	NON_SYNONYMOUS_CODING	275	215	72	A/V	pGc/pTg	Benign
Hspa14	ENSRNOG000000015196	ENSRNOT000000020384	17_85852829_G/C	17	85,852,829	-	G	C	2	0	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	478	478	160	D/H	Gat/Gat	Probably damaging
Hspa14	ENSRNOG000000015196	ENSRNOT000000020384	17_85853753_A/G	17	85,853,753	-	A	G	2	2	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	844	844	282	T/A	Act/Gct	Benign
Hspa14	ENSRNOG000000015196	ENSRNOT000000020384	17_85853981_A/G	17	85,853,981	-	A	G	2	2	0	0	0	6	0	0	NON_SYNONYMOUS_CODING	1072	1072	358	N/D	Asu/Gat	Benign
Dclre1c	ENSRNOG000000015980	ENSRNOT000000021506	rs8173830	17	85,896,829	-	G	A	2	0	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	1922	1867	623	P/S	Cct/Tct	Benign
Dclre1c	ENSRNOG000000015980	ENSRNOT000000021771	rs8173830	17	85,896,829	-	G	A	2	0	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	1922	1867	623	P/S	Cct/Tct	Benign
D444W9_RAT	ENSRNOG000000038045	ENSRNOT000000057895	17_85993589_G/A	17	85,993,589	-	G	A	2	0	2	2	2	0	0	0	NON_SYNONYMOUS_CODING	172	170	57	R/H	cGc/cAt	Benign
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057886	17_86014636_TGTA/-	17	86,014,636	86,014,639	TGTA	-	2	0	2	0	2	0	0	0	FRAMESHIFT_CODING	1063-1066	1063-1066	355-356	-	-	-
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057885	17_86014636_TGTA/-	17	86,014,636	86,014,639	TGTA	-	2	0	2	0	2	0	0	0	FRAMESHIFT_CODING	901-904	901-904	301-302	-	-	-
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057886	17_86014638_TA/-	17	86,014,638	86,014,639	TA	-	2	0	2	2	2	0	0	0	FRAMESHIFT_CODING	1063-1064	1063-1064	355	-	-	-
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057885	17_86014638_TA/-	17	86,014,638	86,014,639	TA	-	2	0	2	2	2	0	0	0	FRAMESHIFT_CODING	901-902	901-902	301	-	-	-
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057886	17_86017004_G/C	17	86,017,004	-	G	C	2	0	2	2	2	0	0	0	NON_SYNONYMOUS_CODING	1007	1007	336	P/R	cCa/cGa	Unknown
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057886	17_86017414_C/A	17	86,017,414	-	C	A	2	0	2	2	2	0	0	0	NON_SYNONYMOUS_CODING	597	597	199	M/I	aG/aT	Unknown
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057885	17_86017414_C/A	17	86,017,414	-	C	A	2	0	2	2	2	0	0	0	NON_SYNONYMOUS_CODING	513	513	171	M/I	aG/aT	Unknown
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057886	17_86017921_-/GT	17	86,017,921	86,017,922	-	GT	0	2	0	0	0	2	2	2	FRAMESHIFT_CODING	104-105	104-105	35	-	-	-
F1M6A7_RAT	ENSRNOG000000038037	ENSRNOT000000057885	17_86017921_-/GT	17	86,017,921	86,017,922	-	GT	0	2	0	0	0	2	2	2	FRAMESHIFT_CODING	6-7	6-7	2-3	-	-	-
Nmt2	ENSRNOG000000026248	ENSRNOT000000030228	17_86060033_C/T	17	86,060,033	-	C	T	2	0	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	134	134	45	S/N	aGc/aAt	Benign
Nmt2	ENSRNOG000000026248	ENSRNOT000000065609	17_86080154_-/G	17	86,080,154	86,080,155	-	G	0	5	2	2	0	2	5	2	FRAMESHIFT_CODING	86-87	86-87	29	-	-	-
Cubn	ENSRNOG000000029047	ENSRNOT000000040052	17_87738747_G/A	17	87,738,747	-	G	A	2	0	0	0	0	6	0	0	NON_SYNONYMOUS_CODING	1310	1310	437	S/L	tGc/tTg	Benign
St8sia6	ENSRNOG000000018171	ENSRNOT000000024475	17_88003467_G/A	17	88,003,467	-	G	A	2	0	0	0	0	0	0	0	NON_SYNONYMOUS_CODING	230	230	77	T/I	aCt/aTt	Benign
Slc39a12	ENSRNOG000000025639	ENSRNOT000000044313	17_88525807_C/A	17	88,525,807	-	C	A	0	2	2	2	2	2	2	2	NON_SYNONYMOUS_CODING,SPICE_SITE	539	539	180	P/Q	cGg/cAg	Benign
Slc39a12	ENSRNOG000000025639	ENSRNOT000000066690	17_88525807_C/A	17	88,525,807	-	C	A	0	2	2	2	2	2	2	2	NON_SYNONYMOUS_CODING,SPICE_SITE	539	539	180	P/Q	cGg/cAg	Benign
Slc39a12	ENSRNOG000000025639	ENSRNOT000000044313	17_88535671_T/G	17	88,535,671	-	T	G	0	2	2	2	2	2	2	2	NON_SYNONYMOUS_CODING	910	910	304	W/G	Tgg/Ggg	Benign
Slc39a12	ENSRNOG000000025639	ENSRNOT000000066690	17_88535671_T/G	17	88,535,671	-	T	G	0	2	2	2	2	2	2	2	NON_SYNONYMOUS_CODING	910	910	304	W/G	Tgg/Ggg	Probably damaging
Slc39a12	ENSRNOG000000025639	ENSRNOT000000044313	17_88575534_T/-	17	88,575,534	88,575,534	T	-	2	0	0	0	0	0	0	0	FRAMESHIFT_CODING	1628	1628	543	-	-	Truncated Protein
Slc39a12	ENSRNOG000000025639	ENSRNOT000000066690	17_88575534_T/-	17	88,575,534	88,575,534	T	-	2	0	0	0	0	0	0	0	FRAMESHIFT_CODING	1520	1520	507	-	-	Truncated Protein

Genotype code: 0 = Homozygous reference allele; 1 = Heterozygous; 2 = Homozygous alternate allele; 6 = Ambiguous. Green square = shared genotype between F344 and susceptible strains; Bold = candidate genes; Blue = most important polymorphism

Extended Data Table 2 | Polymorphism markers for congenic strain genotyping

Rat chr 17	Genetic map		Primers		Expected size (bp)		Genotyping information				
Marker	SHRSP x BN	Physical map possition	FORWARD	REVERSE	WKY	F344	R47A	Sub A	Sub B	Sub C	
D17Rat41	38.27	80234337 - 80234511	CCTTTCCCTTTCCACTCTCC	GGTAAGGGTGGTGGCAGTAG	171	157	WW	WW	WW	WW	
D17Rat44	40.89	81612362 - 81612488	CAGACAAAACCCAGCATTT	AGCAGAAAGAACCAGGCAGA	133	121	WW	WW	WW	WW	
D17Got91		82277435 - 82277611	CCAGACACCAACATCACACC	CCTTCATGTGTGGAGTGTTTATG	160	176	FF	FF	WW	WW	
D17Rat43	40.89	82337230 - 82337376	CACTCACTTGTGGCTGTCT	GAGAAGAAGCTGGAGAGGCA	150	124	FF	FF	WW	WW	
D17Rat42	40.33	82505422 - 82505562	TGCCGCTATTA AAAAGTAAGTGC	CCAAAGGCATAAAATCTTTCC	141	121	FF	FF	WW	WW	
D17Rat62	42.35	83479938 - 83480058	GAAAGGATGGCAGGTTTTTG	TCCACAGGCTCACTGTCACT	143	121	FF	FF	WW	WW	
D17Rat46	42.33	83600152 - 83600282	TGGGTTCTTTCACTCCTTGC	GCTCACCCACACACACATTC	135	125	FF	FF	WW	WW	
D17Rat47	43.34	85072353 - 85072475	CCCTGCTTTCTGCTTTGAAC	TGCATATACGAATTACAGCTCAA	114	126	FF	FF	FF	WW	
Del85862103		85862060 - 85862261	CACCATGAGCTCAGCAGTGT	ACACCTGTGCTCCTCTCAG	203	193	FF	WW	FF	WW	
In85923365		85923262 - 85923495	ACCTTTGGCTCGGTCTCTATC	AAACTTGGGTACCAGGCACCA	235	243	FF	WW	FF	WW	
D17Got93		86032700 - 86032904	CACTACACCTCCCAACGTCC	CTGTTGTGCCTCCTGACTAATG	227	215	FF	WW	FF	WW	
D17Mit8	45.19	87465135 - 87465345	GGTCGGCATTATGGCTAAGA	CTATAGCCTCTAGGGAGGGG	193	195	FF	WW	FF	WW	
D17Rat60	45.19	88268817 - 88269054	GGGGTCCAGCACTTAGCAT	GTTTTGATCATGGGGACGTT	239	241	FF	WW	FF	WW	
D17Rat48	45.19	88667790 - 88667952	CACATGTCTAACTTGCCACATACA	TTTGCTGTTTCTTGTCATGTG	166	156	FF	WW	FF	WW	
Del89391756		89391713 - 89391951	TCCATGTTTTATCACCAGGAAG	ATCTGATGCATGCCATAGCC	230	238	FF	WW	FF	WW	
In90455808		90455697 - 90455921	AAGTTAGCCTTCCCAAGGA	TCTGGTCTTTC CATGTTCC	233	225	FF	WW	FF	WW	
D17Rat131	47.54	93347784 - 933447990	TTAAGAAGGGCAAGCAAGGA	TCCCCATAAAAAGAAAAGGAA	203	213	FF	WW	FF	FF	
D17Rat51	47.54	96587775 - 96587905	TCCCCTGGTCAATCCATTT	ACATGCAGACAGAACTTCTCT	144	148	FF	WW	FF	FF	

Transcriptional control of autophagy–lysosome function drives pancreatic cancer metabolism

Rushika M. Perera^{1,2,3}, Svetlana Stoykova^{1,2*}, Brandon N. Nicolay^{1,3*}, Kenneth N. Ross^{1,2,3}, Julien Fitamant^{1,2,3}, Myriam Boukhali¹, Justine Lengrand^{1,2}, Vikram Deshpande^{3,4}, Martin K. Selig⁴, Cristina R. Ferrone^{1,3,5}, Jeff Settleman¹, Gregory Stephanopoulos⁶, Nicholas J. Dyson^{1,3}, Roberto Zoncu⁷, Sridhar Ramaswamy^{1,2,3}, Wilhelm Haas^{1,3} & Nabeel Bardeesy^{1,2,3}

Activation of cellular stress response pathways to maintain metabolic homeostasis is emerging as a critical growth and survival mechanism in many cancers¹. The pathogenesis of pancreatic ductal adenocarcinoma (PDA) requires high levels of autophagy^{2–4}, a conserved self-degradative process⁵. However, the regulatory circuits that activate autophagy and reprogram PDA cell metabolism are unknown. Here we show that autophagy induction in PDA occurs as part of a broader transcriptional program that coordinates activation of lysosome biogenesis and function, and nutrient scavenging, mediated by the MiT/TFE family of transcription factors. In human PDA cells, the MiT/TFE proteins⁶—MITF, TFE3 and TFEB—are decoupled from regulatory mechanisms that control their cytoplasmic retention. Increased nuclear import in turn drives the expression of a coherent network of genes that induce high levels of lysosomal catabolic function essential for PDA growth. Unbiased global metabolite profiling reveals that MiT/TFE-dependent autophagy–lysosome activation is specifically required to maintain intracellular amino acid pools. These results identify the MiT/TFE proteins as master regulators of metabolic reprogramming in pancreatic cancer and demonstrate that transcriptional activation of clearance pathways converging on the lysosome is a novel hallmark of aggressive malignancy.

Autophagy delivers cargo to lysosomes for degradation, suggesting the possibility that these systems may be coordinately regulated in PDA. Immunostaining for LC3 and LAMP2 revealed significant expansion of both organelles in PDA cell lines compared with non-transformed human pancreatic ductal epithelial cells (HPDEs) (Fig. 1a and Extended Data Fig. 1a). Notably, transmission electron microscopy demonstrated an increase in lysosome number per cell in treatment-naïve PDA specimens relative to normal pancreatic tissue (12.6 ± 4.3 versus 1.0 ± 0.9 ; Fig. 1b). Thus, increased lysosomal biogenesis accompanies the expanded autophagosome compartment in PDA and may facilitate high levels of autophagic flux. Consistent with transcriptional control of these organellar changes, gene set enrichment analysis (GSEA) of multiple independent data sets revealed that human PDA specimens have elevated expression of autophagy–lysosome genes compared with normal pancreatic tissue (Fig. 1c, Extended Data Fig. 1b and Supplementary Tables 1, 2). Accordingly, immunohistochemistry confirmed upregulation of autophagy–lysosome proteins in the tumour epithelium (Fig. 1d).

In normal cells exposed to nutrient stress, the biogenesis of both organelles is under transcriptional regulation by the MiT/TFE subclass of basic helix–loop–helix transcription factors^{6–11}. RNA-sequencing (RNA-seq) data across ten common solid tumour types revealed high relative expression of these factors in PDA, with levels only exceeded in melanoma and kidney cancers, in which MiT/TFE transcription factors are established oncogenes (Extended Data

Fig. 1c). Immunohistochemistry demonstrated overexpression of nuclear-localized TFE3 in the neoplastic epithelium in a subset of PDA (staining scores ≥ 2 in 23% of PDA versus 3% of normal pancreas specimens; $P < 0.001$) (Fig. 1e and Extended Data Fig. 1d; see Methods). Similarly, microdissected specimens and xenografts exhibited frequent upregulation of *TFEB* and *MITF* messenger RNA in PDA cells relative to normal ductal epithelium, and subsets of PDA cell lines showed MiT/TFE overexpression compared with HPDE cells (Extended Data Fig. 1e–h). Generally, a single MiT/TFE family member predominated in individual specimens.

GSEA of human primary PDA data sets and cultured PDA cell lines showed strong correlation between expression of MiT/TFE factors and the autophagy–lysosome signature (Fig. 1f and Extended Data Fig. 1i–k). Accordingly, knockdown of TFE3 in the 898T PDA cell line (TFE3-high, TFEB/MITF-low) resulted in prominent repression of this signature (global RNA-seq; Fig. 1g, h). Chromatin immunoprecipitation (ChIP) confirmed that MITF and TFE3 bound to multiple autophagy and lysosome genes bearing a consensus coordinated lysosomal expression and regulation (CLEAR) element^{7,8} in PDA cells (Extended Data Fig. 2a). Moreover, knockdown of MITF, TFE3 or TFEB caused downregulation of numerous CLEAR-bearing genes in a series of PDA cell lines with high relative expression of that MiT/TFE family member, whereas no significant changes were seen in non-transformed pancreatic lines (HPDE and HPNE) or in a pancreatic neuroendocrine tumour cell line (QGP1) (Extended Data Fig. 2b–d). Expression of RNA interference (RNAi)-resistant *MITF* or *TFE3* complementary DNA restored target gene expression whereas dominant-negative MITF recapitulated the effects seen with RNAi (Extended Data Fig. 2e, f). Thus, MiT/TFE proteins act selectively in PDA cells to regulate a broad autophagy–lysosome program under basal conditions.

In non-transformed cells grown in nutrient-replete conditions, the MiT/TFE proteins are phosphorylated by mTORC1 at the lysosome membrane, leading to their interaction with 14-3-3 proteins and cytoplasmic retention, whereas mTORC1 inactivation upon starvation enables their nuclear translocation^{9–11}. Correspondingly, HPDE and HPNE cells exhibited predominantly cytoplasmic residence of endogenous and ectopically expressed TFE3, MITF and TFEB under full nutrient conditions, and showed nuclear translocation after starvation or treatment with the mTOR inhibitor Torin 1 (Fig. 2a–c and Extended Data Fig. 3a). In stark contrast, a series of PDA cell lines showed constitutive nuclear localization of each MiT/TFE protein, regardless of nutrient status or treatment with Torin 1 or with inhibitors of MEK, another pathway implicated in their regulation⁸ (Fig. 2a–c and Extended Data Fig. 3b–i). Immunoblot for phospho-p70S6K, a readout of mTORC1 activity, and immunostaining for mTOR and LAMP2, indicated that mTORC1 was active and

¹Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ³Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁴Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁵Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁶Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁷Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

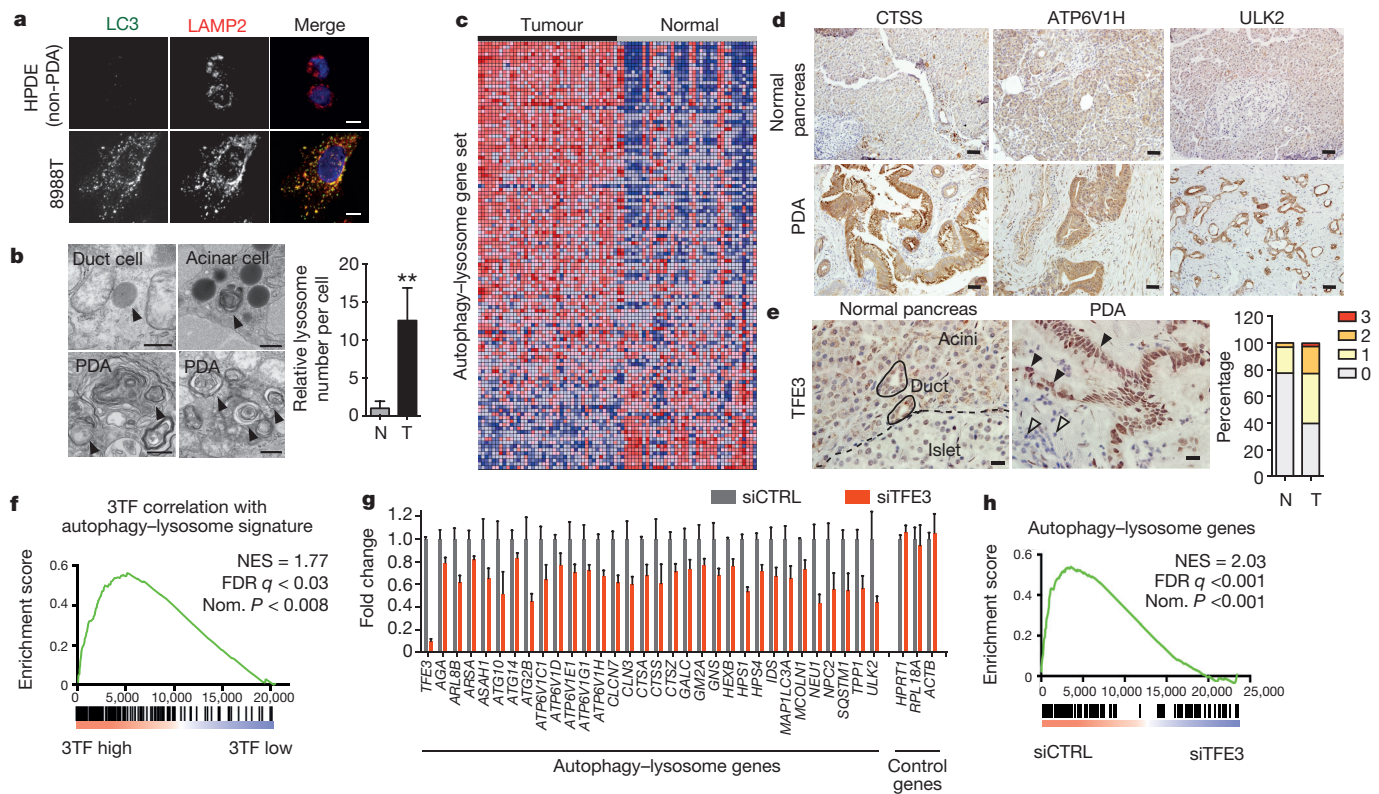


Figure 1 | Coordinate induction of an autophagy-lysosome gene program in PDA by MiT/TFE proteins. **a**, Immunofluorescence staining showing extensive overlap of autophagosomes (LC3) and lysosomes (LAMP2) in 8988T cells compared with HPDE cells. **b**, Representative transmission electron micrographs showing increased abundance of lysosomes in PDA compared with normal pancreas. Relative lysosome numbers per cell are quantified (see Methods). $N = 473$ cells from 4 normal specimens and 406 cells from 3 PDA specimens. $^{**}P < 0.001$. N, normal; T, tumour. **c**, Upregulation of autophagy-lysosome genes in PDA relative to matched normal tissue (see Supplementary Table 2). **d, e**, Immunohistochemistry showing upregulation of autophagy and lysosomal proteins (**d**) and nuclear-localized TFE3 (**e**) in the PDA epithelium (filled arrowheads) compared with normal pancreas or

stromal cells (open arrowheads). **e**, Graph shows quantification of TFE3 staining intensity (0 = no staining to 3 = high staining) in normal ($N = 31$) and PDA samples ($N = 354$). **f**, GSEA showing correlation between MiT/TFE (3TF) expression and the autophagy-lysosome gene signature in primary human PDA. FDR, false discovery rate; NES, normalized enrichment score; Nom. $P < 0.008$. **g, h**, TFE3 knockdown in 8988T cells causes coordinated downregulation of autophagy-lysosome genes. **g**, RNA-seq values from $N = 3$ independent experiments; Error bars represent mean \pm standard deviation (s.d.). $P < 0.05$ for each gene. **h**, GSEA. Scale bars: 11 μm (**a**); 500 nm (**b**); 50 μm (**d**); 20 μm (**e**). Significance was analysed using two-tailed Student's t -test.

demonstrated amino-acid-regulated lysosomal association in PDA cells (Extended Data Fig. 3j, k). In addition, all cells exhibited co-immunoprecipitation of TFE3 and MITF with 14-3-3 and loss of binding upon Torin 1 treatment, although the fractional binding in PDA cells was lower, consistent with the predominantly nuclear residence of the MiT/TFE proteins (Extended Data Fig. 3l). Thus, PDA cells show constitutive MiT/TFE activation despite displaying intact mTORC1 signalling.

To uncover machinery that overrides cytoplasmic retention mechanisms, we employed affinity purification and quantitative proteomics¹² to identify PDA-specific TFE3-interacting proteins. Notably, importin 8 (IPO8), a member of the importin- β family of nucleocytoplasmic transporters^{13,14}, was significantly enriched in TFE3 immunoprecipitates in PDA cells but not in HPDE cells (Extended Data Fig. 4a). We confirmed that Flag-tagged TFE3 bound endogenous IPO8 specifically in PDA cells (Fig. 2d and Extended Data Fig. 4b). In addition, IPO8 was expressed at elevated levels in human PDA specimens and cell lines compared with controls (Fig. 2e and Extended Data Fig. 4c, d).

IPO8 and its most closely related homologue, IPO7, direct nuclear import of specific cargo, including growth regulatory proteins^{14–19}. Accordingly, IPO8 knockdown caused a marked decrease in nuclear TFE3 and reduced overall TFE3 protein levels in multiple PDA cell lines (Fig. 2f, left, and Extended Data Fig. 4e). Similar effects were observed for both MITF and TFEB upon combined IPO8 and

IPO7 knockdown (Fig. 2f, right, bottom and Extended Data Fig. 4f). Importantly, in non-transformed HPDE and HPNE cells, IPO8 knockdown did not affect basal (cytoplasmic) TFE3 levels or Torin-1-induced nuclear translocation of TFE3 (Extended Data Fig. 4g, h). Moreover, MiT/TFE mRNA levels were not significantly altered by IPO7/IPO8 inactivation (Extended Data Fig. 4i). These findings suggest that importins may act selectively in PDA to regulate the stability of MiT/TFE proteins in addition to facilitating their nuclear transport. Indeed, cycloheximide treatment studies revealed that IPO8 knockdown accelerated TFE3 turnover in 8988T cells (Extended Data Fig. 4j). Thus, IPO8-dependent nuclear accumulation of MiT/TFE factors results in their stabilization and upregulation of their transcriptional programs in PDA cells regardless of nutrient status.

We found that constitutive activation of MiT/TFE proteins was critical for autophagy-lysosome function in PDA cells. Depletion of MiT/TFE proteins across a series of PDA cell lines resulted in striking defects in lysosome morphology and increased lysosome diameter, (986.4 ± 30.7 nm versus $2,722 \pm 72.6$ nm for control siRNA (siCTRL) and siTFE3, respectively), effects commonly associated with lysosomal stress and defective proteolysis^{20,21} (Fig. 2g and Extended Data Fig. 5a), whereas HPDE cells were minimally affected (Extended Data Fig. 5b). These responses were phenocopied by treatment of PDA cells with the vacuolar-type H^+ -ATPase inhibitor Bafilomycin A1 (BafA1) (Fig. 2g). Correspondingly, transmission electron microscopy and

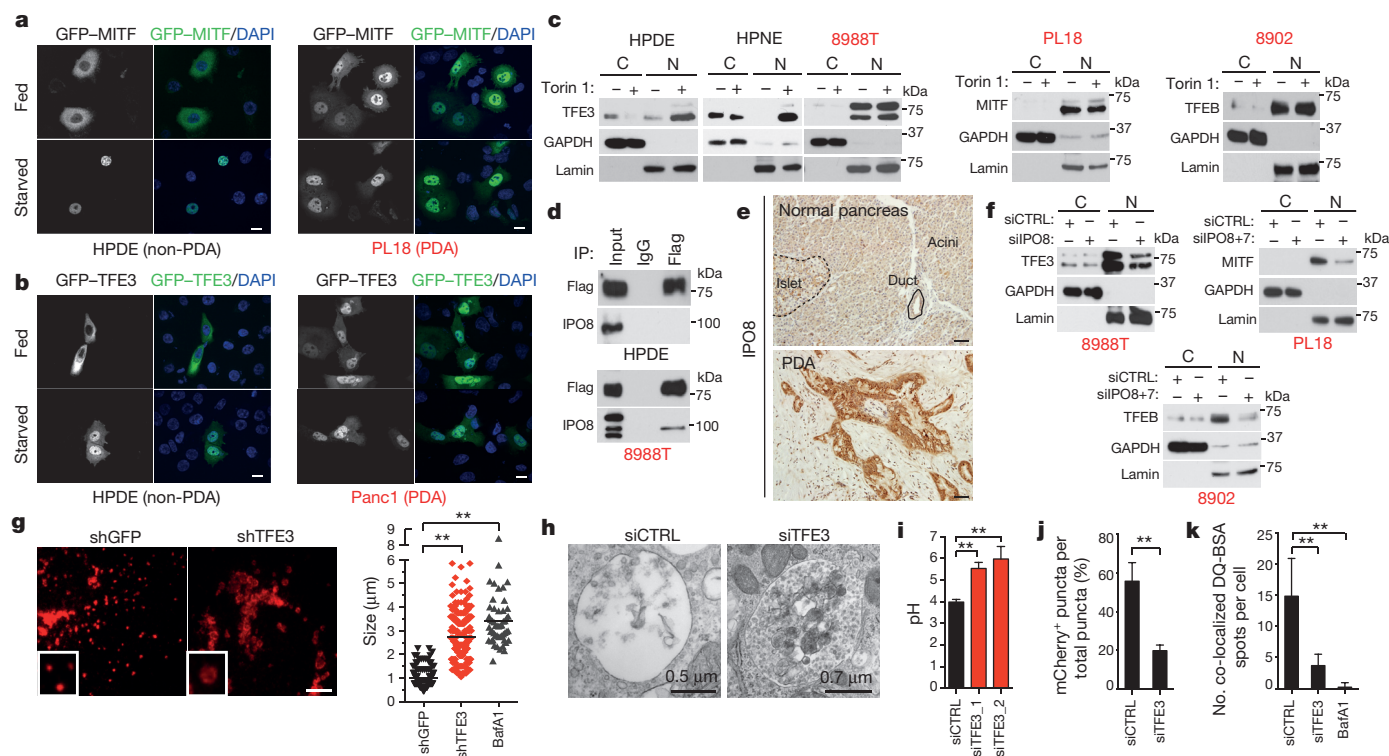


Figure 2 | Constitutive nuclear import of MiT/TFE factors controls autophagy-lysosome function in PDA. **a, b,** Fluorescence microscopy showing localization of GFP-MITF (**a**) and GFP-TFE3 (**b**) in the indicated cells under fed and starved (3 h Hank's balanced salt solution (HBSS)) conditions. DAPI, 4',6-diamidino-2-phenylindole. **c,** Immunoblots for TFE3 (left), MITF (middle) and TFE3 (right) in cytoplasmic (C) and nuclear (N) fractions of cells treated with vehicle or Torin 1 (250 nM) for 1 h. **d,** Flag-TFE3 was stably expressed in cells and lysates were immunoprecipitated with anti-Flag antibody or control IgG and immunoblotted for Flag (TFE3) or IPO8. **e,** Representative IPO8 immunostaining in human specimens. **f,** PDA cell lines were transfected with the indicated siRNAs, fractionated and immunoblotted for TFE3 (898T cells; left), MITF (PL18 cells; right) or TFE3 (8902 cells; bottom). **g,** TFE3 knockdown causes aberrant lysosomal morphology and increased size as shown by LAMP2 staining. Inset: magnified view. Graph (right): quantification of lysosome diameter in cells expressing shGFP ($N = 151$), shTFE3 ($N = 193$),

or treated with BafA1 ($N = 52$). $^{**}P < 0.0001$; bar, mean. **h,** Electron micrographs showing accumulation of undigested cargo in lysosomes upon TFE3 knockdown. **i,** Measurement of lysosomal pH in 8988T cells transfected with the indicated siRNAs. $N = 3$ independent experiments; $^{**}P < 0.001$ (see Methods). **j,** Quantification of total number of autolysosomes (mCherry⁺/GFP⁺ spots) from $N = 10$ cells per condition using a tandem mCherry-GFP-LC3 reporter. Error bars represent mean \pm standard error of the mean (s.e.m.). $^{**}P < 0.0001$. **k,** Proteolysis of macropinocytosed protein is impaired by TFE3 knockdown or BafA1 treatment as determined by pulse-chase with DQ-BSA. Degradation of DQ-BSA in lysosomes is quantified (number of fluorescent spots per cell co-localizing with LAMP2⁺ lysosomes for $N = 3$ independent experiments with at least 50 cells scored per experiment). $^{**}P < 0.0001$. Scale bars: 18 μ m (**a, b**); 50 μ m (**e**); 7.5 μ m (**g**). Significance was analysed using two-tailed Student's *t*-test.

immunofluorescence microscopy of TFE3-knockdown cells revealed accumulation of autolysosomes containing undigested cargo (Fig. 2h and Extended Data Fig. 5c, d). Moreover, there was a pronounced increase in lysosomal pH as demonstrated using the pH-sensitive dye Oregon Green 514 coupled to dextran (Fig. 2i). In addition, mCherry-GFP-LC3 autophagy reporter assays revealed that TFE3 inactivation decreased autophagic flux in 8988T cells ($55.7 \pm 9.6\%$ mCherry⁺/GFP⁺ spots in siCTRL versus $19.8 \pm 2.9\%$ in siTFE3; Fig. 2j). Finally, to assay proteolytic activity, 8988T cells were fed boron-dipyrromethene (BODIPY)-dye-conjugated bovine serum albumin (BSA; DQ-BSA), which is taken up by macropinocytosis and fluoresces after proteolysis in lysosomes. TFE3 knockdown or BafA1 treatment resulted in dramatic decreases in fluorescent puncta co-localized with LAMP2 (14.8 ± 6.1 spots per cell in siCTRL versus 3.7 ± 1.8 and 0.2 ± 0.7 spots per cell in siTFE3 and BafA1, respectively; Fig. 2k). Tetramethylrhodamine (TMR)-dextran uptake was not affected, indicating specific impairment in lysosomal proteolysis rather than in internalization of extracellular cargo (data not shown). Notably, lysosomal breakdown of extracellular albumin scavenged through macropinocytosis is an important nutrient source in PDA²². Thus, by governing both autophagic flux and lysosomal catabolism, the MiT/TFE proteins support an integrated cellular clearance program that enables efficient processing

of cargo from autophagy as well as macropinocytosis, providing PDA cells with access to critical sources of both intracellular and extracellular nutrients.

MITF or TFE3 overexpression in pancreatic cells reinforced these findings, demonstrating induction of autophagy-lysosome genes, LC3B foci and lipidated LC3-II that was further enhanced after treatment with chloroquine, an inhibitor of lysosome acidification, indicating marked augmentation of autophagic flux (Extended Data Fig. 6a–e).

Cargo degraded in the lysosome generates metabolic intermediates that may feed into multiple pathways^{23,24}. To obtain insight into the metabolic outputs of MiT/TFE-controlled autophagy-lysosome function, we conducted global metabolite profiling of 8988T and PSN1 cells transfected with TFE3-targeted siRNAs. These studies revealed statistically significant changes common to both cell lines in 15.2% (53/347) of detected metabolites. Most prominently altered were amino acids and their breakdown products, with 31% (25/80) showing decreased abundance, whereas only restricted changes were observed in other general metabolite groupings (Fig. 3a and Supplementary Table 3). TFE3 silencing did not affect amino acid uptake (Extended Data Fig. 7a), suggesting that the autophagy-lysosome system may supply a considerable fraction of intracellular amino acids irrespective of their external availability in PDA. In support of this model, TFE3

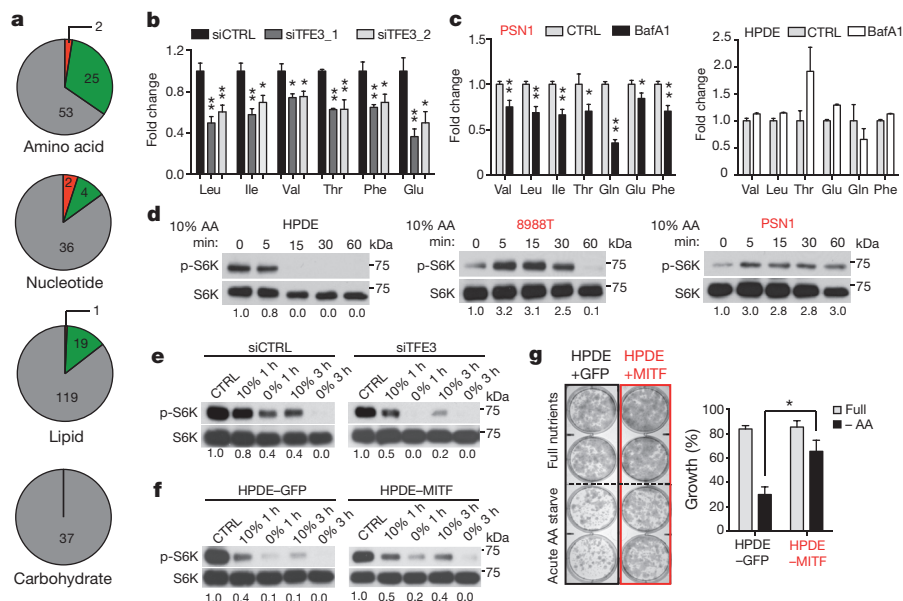


Figure 3 | MiT/TFE proteins maintain autolysosome-derived pools of amino acids. **a**, Global metabolite profiling of 8988T and PSN1 cells after TFE3 knockdown reveals a preferential decrease in amino acids compared with other general metabolite groups (see Methods). Grey, no change; red, increase; green, decrease. **b**, Quantification of fold change in amino acid levels in 8988T cells transfected with siCTRL or siTFE3. * $P < 0.03$, ** $P < 0.003$. **c**, BafA1 causes a decrease in amino acid levels in PDA cells (left), while amino acids are unchanged or increased in HPDE cells (right). * $P < 0.05$, ** $P < 0.005$. **d**, Immunoblot for p-p70S6K levels in the indicated cells at sequential times

after 10% amino acid (AA) starvation. **e**, p-p70S6K immunoblots in 8988T cells transfected with siCTRL or siTFE3 and grown under control conditions or in 0% or 10% amino acids for the indicated times. **f**, **g**, HPDE-MITF cells show sustained p-p70S6K levels under amino acid starvation (**f**) and enhanced colony formation under transient starvation with 10% amino acids (quantified in the graph) (**g**). * $P < 0.001$. For all graphs, error bars indicate mean \pm standard deviation (s.d.) for $N = 3$ independent experiments. Significance was analysed using two-tailed Student's *t*-test.

knockdown, BafA1 treatment, and inactivation of autophagy by ATG5 knockdown significantly decreased intracellular amino acid levels in PDA cell lines but not in control cells (Fig. 3b, c and Extended Data Fig. 7b–d), and led to activation of AMP-activated protein kinase (AMPK) and decreased cellular ATP levels, suggesting induction of energy stress (Extended Data Fig. 8a–c). Reciprocally, overexpression of MITF or TFE3 in HPDE cells caused downregulation of AMPK signalling (Extended Data Fig. 8d).

These findings imply that PDA cells should have increased capacity to buffer amino acid levels under conditions of nutrient deprivation. To test this, we switched cells to low amino acid concentration media (10% of normal) and monitored phosphorylation of p70S6K, a readout of mTORC1 activity that is sensitive to fluctuations in intracellular amino acid levels²⁵. HPDE cells showed extinction of phosphorylated (p)-p70S6K within 5–15 min in 10% amino acid media, whereas PDA cells (8988T and PSN1) maintained robust levels for >30 min (Fig. 3d). Correspondingly, TFE3 inactivation in 8988T cells greatly accelerated the decline in p-p70S6K (Fig. 3e). Reciprocally, MITF overexpression in HPDE cells caused p-p70S6K levels to be sustained upon 10% amino acid starvation, and markedly increased clonogenic growth in this setting while not affecting clonogenicity in full nutrients (Fig. 3f, g). Taken together, these data indicate that PDA cells rely on autophagy–lysosome function for maintenance of intracellular amino acid pools, which facilitates survival in response to nutrient stress.

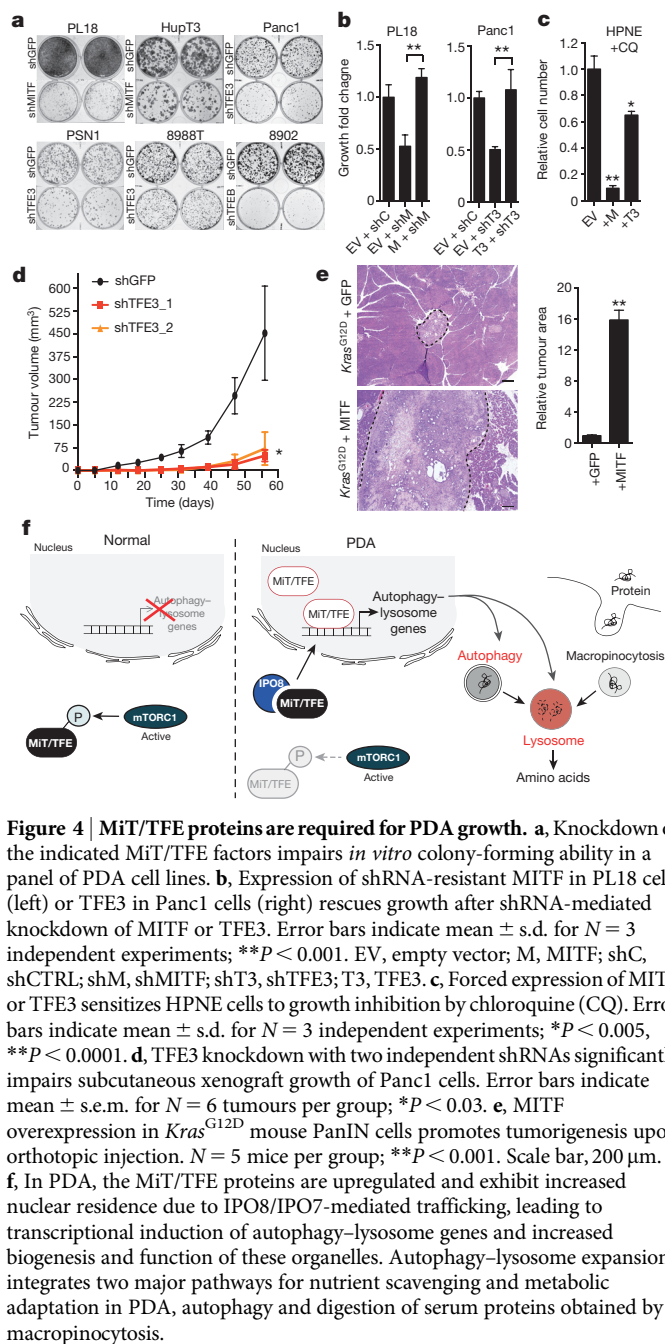
Consistent with key roles of MiT/TFE proteins in organelle function and metabolic regulation, the growth of PDA cell lines expressing high endogenous levels of MITF, TFE3 or TFEB was significantly impaired by knockdown of the relevant factor, an effect rescued by co-expression of short hairpin RNA (shRNA)-resistant cDNAs (Fig. 4a, b and Extended Data Fig. 9a). By contrast, the growth of HPNE and QGP1 cells was unaffected (Extended Data Fig. 9a). As previously reported, PDA cells were broadly sensitive to chloroquine treatment compared with control cells² (Extended Data Fig. 9b). Moreover, ectopic

expression of MITF or TFE3 rendered HPNE cells hypersensitive to chloroquine, linking MiT/TFE-regulated clearance pathways to these growth phenotypes (Fig. 4c).

To extend our findings in primary patient-derived samples, we examined a series of early passage PDA cultures. Importantly, these cells also showed high basal autophagy, nuclear-localized MiT/TFE proteins, and MiT/TFE-dependent autophagy–lysosome gene expression, organelle function, and colony-forming ability (Extended Data Fig. 9c–g).

Notably, knockdown of TFE3 and MITF virtually abolished xenograft tumour growth of Panc1 and 8988T cells, and PL18 cells, respectively (Fig. 4d and Extended Data Fig. 9h). Conversely, MITF overexpression enhanced the tumorigenicity of primary *Kras*^{G12D}-expressing mouse pancreatic epithelial cells. Whereas *Kras*^{G12D} control cells formed only focal low-grade pancreatic intraepithelial neoplasia (PanIN)-like lesions by 6 weeks after orthotopic injection, co-expression of MITF resulted in large invasive tumours in 3/5 mice (Fig. 4e and Extended Data Fig. 9i). Thus, in keeping with their requirement for augmenting autophagy–lysosome function, MiT/TFE proteins are potent drivers of PDA pathogenesis *in vivo*.

Our work places a new focus on lysosome regulation by MiT/TFE proteins as a nexus for metabolic reprogramming in PDA cells. Increased lysosome activity integrates major routes for nutrient scavenging—autophagy and macropinocytosis—thereby maintaining intracellular amino acid availability (Fig. 4f). Escape of MiT/TFE factors from inhibition by mTORC1 enables cancer cells to maintain robust activation of anabolic pathways while simultaneously benefiting from the metabolic fine-tuning and adaptation to stress afforded by activation of autophagy and lysosomal catabolism. Thus, we propose that lysosome activation under the control of the MiT/TFE transcriptional program is a novel hallmark of PDA. These findings, together with recent work^{26–28}, suggest that targeting aberrant lysosomal function has a potential therapeutic benefit in cancer.



Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 March 2014; accepted 20 May 2015.

Published online 13 July 2015.

- White, E. Exploiting the bad eating habits of Ras-driven cancers. *Genes Dev.* **27**, 2065–2071 (2013).
- Yang, S. *et al.* Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25**, 717–729 (2011).
- Rosenfeldt, M. T. *et al.* p53 status determines the role of autophagy in pancreatic tumour development. *Nature* **504**, 296–300 (2013).
- Yang, A. *et al.* Autophagy is critical for pancreatic tumor growth and progression in tumors with p53 alterations. *Cancer Discov.* **4**, 905–913 (2014).

- Kroemer, G., Marino, G. & Levine, B. Autophagy and the integrated stress response. *Mol. Cell* **40**, 280–293 (2010).
- Haq, R. & Fisher, D. E. Biology and clinical relevance of the microphthalmia family of transcription factors in human cancer. *J. Clin. Oncol.* **29**, 3474–3482 (2011).
- Settembre, C. *et al.* TFEB links autophagy to lysosomal biogenesis. *Science* **332**, 1429–1433 (2011).
- Sardiello, M. *et al.* A gene network regulating lysosomal biogenesis and function. *Science* **325**, 473–477 (2009).
- Settembre, C. *et al.* A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. *EMBO J.* **31**, 1095–1108 (2012).
- Roczniak-Ferguson, A. *et al.* The transcription factor TFEB links mTORC1 signaling to transcriptional control of lysosome homeostasis. *Sci. Signal.* **5**, ra42 (2012).
- Martina, J. A. *et al.* The nutrient-responsive transcription factor TFE3 promotes autophagy, lysosomal biogenesis, and clearance of cellular debris. *Sci. Signal.* **7**, ra9 (2014).
- Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8**, 937–940 (2011).
- Raices, M. & D'Angelo, M. A. Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nature Rev. Mol. Cell Biol.*
- Chook, Y. M. & Suel, K. E. Nuclear import by karyopherin- β s: recognition and inhibition. *Biochim. Biophys. Acta* **1813**, 1593–1606 (2011).
- Zehorai, E. & Seger, R. Beta-like importins mediate the nuclear translocation of mitogen-activated protein kinases. *Mol. Cell Biol.* **34**, 259–270 (2014).
- Yao, X., Chen, X., Cottonham, C. & Xu, L. Preferential utilization of Imp7/8 in nuclear import of Smads. *J. Biol. Chem.* **283**, 22867–22874 (2008).
- Waldmann, I., Walde, S. & Kehlenbach, R. H. Nuclear import of c-Jun is mediated by multiple transport receptors. *J. Biol. Chem.* **282**, 27685–27692 (2007).
- Xu, L. *et al.* Msk is required for nuclear import of TGF- β /BMP-activated Smads. *J. Cell Biol.* **178**, 981–994 (2007).
- Golomb, L. *et al.* Importin 7 and exportin 1 link c-Myc and p53 to regulation of ribosomal biogenesis. *Mol. Cell* **45**, 222–232 (2012).
- Ballabio, A. & Gieselmann, V. Lysosomal disorders: from storage to cellular damage. *Biochim. Biophys. Acta* **1793**, 684–696 (2009).
- Settembre, C., Fraldi, A., Medina, D. L. & Ballabio, A. Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nature Rev. Mol. Cell Biol.* **14**, 283–296 (2013).
- Commissio, C. *et al.* Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497**, 633–637 (2013).
- Rabinowitz, J. D. & White, E. Autophagy and metabolism. *Science* **330**, 1344–1348 (2010).
- Singh, R. *et al.* Autophagy regulates lipid metabolism. *Nature* **458**, 1131–1135 (2009).
- Zoncu, R., Efeyan, A. & Sabatini, D. M. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nature Rev. Mol. Cell Biol.* **12**, 21–35 (2011).
- Petersen, N. H. *et al.* Transformation-associated changes in sphingolipid metabolism sensitize cells to lysosomal cell death induced by inhibitors of acid sphingomyelinase. *Cancer Cell* **24**, 379–393 (2013).
- Saftig, P. & Sandhoff, K. Cancer: killing from the inside. *Nature* **502**, 312–313 (2013).
- Narita, M. *et al.* Spatial coupling of mTOR and autophagy augments secretory phenotypes. *Science* **332**, 966–970 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank A. Kimmelman, L. Ellisen, W. Kim and R. Mostoslavsky for advice and helpful comments on the manuscript, S. Gygi for access to proteomics data analysis software, F. Kottakis, Y. Mizukami and M. Leisa for technical support, and C. Ivan for bioinformatics support. This work was supported by grants from the National Institutes of Health (P50CA127003, P01 CA117969-07, R01 CA133557-05) and the Linda J. Verville Cancer Research Foundation to N.B. N.B. holds the Gallagher Endowed Chair in Gastrointestinal Cancer Research. R.M.P. holds a Hirshberg Foundation for Pancreatic Cancer seed grant. N.B. and R.M.P. are members of the Andrew Warshaw Institute for Pancreatic Cancer Research.

Author Contributions R.M.P. and N.B. conceived and designed the study. R.M.P. and S.S. performed all experiments involving PDA cells and mouse models. R.M.P. and B.N.N. performed the metabolite measurements. J.F. and J.L. performed immunohistochemistry on human tissue sections. M.B. and W.H. performed quantitative proteomics measurements and analysis. K.N.R. and S.R. performed computational analysis. V.D. and M.K.S. performed pathology assessment and electron microscopy analysis. C.R.F. provided essential reagents. N.J.D. and G.S. supervised the metabolite analysis. R.Z. and J.S. contributed to the study design. R.M.P. and N.B. wrote the manuscript with feedback from all authors.

Author Information RNA-sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE62077. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.B. (Bardeesy.Nabeel@MGH.harvard.edu).

METHODS

Materials. Reagents were obtained from the following sources: antibodies against MITF (sc-71587) from Santa Cruz Biotechnology; GAPDH (MAB374) from Millipore; TFE3 from Cell Marque (MRQ-37; 354R-14); LAMP2 (ab-25631) from Abcam; TFE3 (4240), LC3B (3868), phospho-T389 S6K1 (9234), S6K1 (2708), phospho-AMPK (Thr 172) (2535), AMPK (2603), phospho-ACC (Ser 79) (3661), ACC (3676), Lamin (2032), mTOR (2893) and the Flag epitope (2368) from Cell Signaling Technology; IPO8 (NBP2-24751) and LC3-II (NB600-1384) from Novus Biologicals; CTSS (HPA002988), ATP6V1H (HPA023421) and ULK2 (HPA009027) from Sigma-Aldrich; Flag M2 affinity gel, amino acids from Sigma Aldrich; RPMI, DMEM, fetal bovine serum (FBS) and dialysed FBS (dFBS), Alexa 488- and 568-conjugated secondary antibodies; 70 kDa dextran-Oregon Green 514, DQ-Green-BSA from Invitrogen; amino-acid-free RPMI from US Biologicals; Bafilomycin A1 from Tocris.

Cell culture. Cell lines were obtained from the American Type Culture Collection. Patient-derived PDA cultures were generated at Massachusetts General Hospital (MGH). Informed consent was obtained from all subjects. STR profiling was performed for all cell lines by the MGH Center for Molecular Therapeutics. Primary human samples and PDA cell lines were generated from ascites fluid under Institutional Review Board (IRB)-approved protocols 02-240 and 2007P001918 (see also Supplementary Table 4). All lines used had verified activating KRAS mutations. Cells were cultured in the following media: 8988T, Panc1, PL18, PSN1, 8902, in DMEM supplemented with 10% FBS; HupT3, KLM1 in RPMI with 10% FBS; HPDE cells were cultured as described previously²⁷. Negative mycoplasma contamination status of all cell lines and primary cells used in the study was established using LookOut Mycoplasma PCR Kit (Sigma, MP0035).

Constructs. N-terminal Flag-tagged human MITF-H isoform expression constructs were generated by subcloning the cDNA of MITF-H (Origene) into the NheI and EcoRI sites of the pLJM1 (Addgene) lentiviral vector, the NotI and EcoRI sites of the pRetroX-Tight-Pur (Clontech) doxycycline-inducible vector or the EcoRI and SalI sites of the pBabe-puro retroviral vector. Dominant-negative MITF-H isoform lacking the 5' transactivation domain was generated as previously described for the MITF-M isoform³⁰. Briefly, the first 280 amino acids corresponding to 840 bp was deleted followed by deletion of R301 (corresponding to R218 of MITF-M). This cDNA was cloned into the NheI and EcoRI sites of pLJM1. N-terminal Flag-tagged full-length human TFE3 expression constructs were generated by subcloning the cDNA of TFE3 from pcDNA3.1-(HA)₂-TFE3 (obtained from D. Fisher) into the BamHI and EcoRI sites of the pBabe-puro retroviral vector. GFP-tagged MITF and TFE3 were generated by subcloning the cDNA of MITF-H into BglII and SalI sites and TFE3 into HindIII and BamHI sites of the pEGFP-C1 (Clontech) vector. siRNA-resistant expression constructs of MITF and TFE3 were generated via QuikChange Site-Directed Mutagenesis (Agilent Technologies). mCherry-GFP-LC3 was a gift from A. Kimmelman.

siRNA and lentiviral-mediated shRNA targets. siRNAs against MITF, TFE3, TFE3, ATG5 and IPO8 were purchased from Ambion and Thermo Scientific. The RNAi Consortium clone IDs for the shRNAs used in this study are as follows: shMITF, TRCN0000329793 (CGGGAACCTTGATTGATCTTT), TRCN0000019122 (CGTGACTATATCCGAAGATT); shTFE3, TRCN0000013108 (CCCACTTGGTGCTAATAGCT), TRCN0000013109 (CGATGTCCTTGGCTACATCA). shRNA constructs against human TFE3 were generated by annealing oligonucleotides against human TFE3 (corresponding to the sequence of the siRNA described later) into the AgeI and EcoRI sites of pLKO.1-puro vector. TFE3 target sequences are the following: GCAGCTCCGAATCCAGGAAGCT; CGCAGGCGATTCAACATTAAC; GGAATCTGCTTGATGTGTACA.

Cell proliferation and colony formation assay. Cells were plated in 12-well plates at 30,000 cells per well in 2 ml of complete culture media or in 6-well plates at 2,000 cells per well in 2 ml of media. At the indicated time points, cells were trypsinized and counted. Colony plates were fixed 7 days post-plating in 4% paraformaldehyde and stained with 0.1% crystal violet.

Amino acid starvation and BafA1 treatment. Ten per cent amino acid (AA) starvation was conducted by incubating cells in AA-free RPMI supplemented with 10 mM glucose, and 10% AA calculated relative to levels present in RPMI media. Stock solutions of AA were made from individual powders. Where BafA1 treatment was performed, cells were incubated with 150 nM BafA1. Acute AA starvation during colony formation was conducted by incubating cells in AA-free media for 18 h on day 2 post-plating. Media was then changed to complete growth media for an additional 6 days of cell growth.

SDS-PAGE analysis. Cells were lysed in ice-cold lysis buffer (150 mM NaCl, 20 mM Tris (pH 7.5), 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 2.5 mM sodium pyrophosphate, 1 mM β -glycerophosphate, 1 mM sodium vanadate, and one tablet of EDTA-free protease inhibitors (Roche) per 10 ml). Samples were clarified by centrifugation and protein content measured using BCA protein assay kit (Thermo Scientific). Thirty milligrams of protein was resolved on 9%

SDS-PAGE gels and transferred onto PVDF membranes (GE Healthcare Life Sciences). Membranes were blocked in Tris-buffered saline (TBS) containing 5% non-fat dry milk and 0.1% Tween 20 (TBS-T), before incubation with primary antibody overnight at 4 °C. The membranes were then washed with TBS-T followed by exposure to the appropriate horseradish peroxidase (HRP)-conjugated secondary antibody for 45 min and visualized on Kodak X-ray film using the enhanced chemiluminescence (ECL) detection system (Thermo Scientific). Subcellular fractionation was performed using the Thermo Scientific NE-PER nuclear and cytoplasmic fractionation kit according to the manufacturer's procedures. For immunoprecipitation experiments, 1–3 mg of clarified protein lysate from cell lines stably expressing Flag-tagged MITF/TFE proteins was captured on anti-Flag antibody-conjugated protein A-agarose beads (M2-Flag affinity gel; Sigma-Aldrich). Immune complexes were washed and analysed via SDS-PAGE and western blotting. Immunoglobulin G antibody was used as a control.

Metabolite extraction. Samples were extracted while cells were in the exponential growth phase. Wells were washed with 1× PBS and aspirated. Six-hundred microlitres of −80 °C MeOH:H₂O (50:50) was added to each well and the plate was then placed at −80 °C for 1–1.5 h. As plates were thawing on ice, cells were detached by scraping and transferred to a 1.5 ml tube and placed on dry ice. Samples were vortexed at 4 °C for 30 s. To clarify homogenates, samples were centrifuged at 20,000g, for 7 min at 4 °C. Clarified supernatants were transferred to new 1.5 ml tubes. Six-hundred microlitres of chloroform (stabilized with amylene) was added to each sample, one at a time, and the sample was immediately vortexed at room temperature for 30 s, and then placed on ice. Phase separation was achieved by centrifugation at 20,000g, for 0.25 h at 4 °C. The aqueous (polar) phase was extracted and transferred to a fresh 1.5 ml tube. The samples were evaporated by SpeedVac, snap frozen in liquid nitrogen, and stored at −80 °C for further processing.

Derivatization (all samples). Evaporated samples were warmed to room temperature by quickly spinning them in a SpeedVac. Samples were dissolved in 30 μ l of 2% methoxyamine hydrochloride in pyridine (MOX) (Pierce) at 37 °C for 1.5 h. Samples were derivatized by adding 45 μ l of *N*-methyl-*N*-(tert-butyl)-dimethylsilyl)trifluoroacetamide (MBTSTFA) + 1% tert-butyltrimethylchlorosilane (TBDMS; Pierce) at 60 °C for 1 h.

Gas chromatography-tandem mass spectrometry analysis. Gas chromatography-mass spectrometry (GC-MS) analysis was performed as described^{31,32}. Briefly, analysis was performed on an Agilent 6890 GC instrument that contained a 30 m DB-35MS capillary column, which was interfaced to an Agilent 5975B MS. Electron impact (EI) ionization was set at 70 eV. Each analysis was operated in scanning mode, recording mass-to-charge-ratio spectra in the range of 100–605 *m/z*. For each sample, 1 μ l was injected at 270 °C, using helium as the carrier gas at a flow rate of 1 ml min^{−1}. To mobilize metabolites, the GC oven temperature was held at 100 °C for 3 min and increased to 300 °C at 3.5 °C min^{−1}. Samples were analysed in triplicate. Metabolite levels were normalized to cell number and/or protein and plotted as fold change relative to control conditions.

Immunofluorescence assays. Cells were plated on fibronectin-coated glass coverslips at 100,000–300,000 cells per coverslip. Twelve-to-sixteen hours later, the slides were rinsed with PBS once and fixed for 15 min with 4% paraformaldehyde at room temperature or for 5 min with −20 °C methanol. The slides were rinsed twice with PBS and cells were permeabilized with 0.05% Triton X-100 for 2 min. After rinsing twice with PBS, the slides were incubated with primary antibody in 5% normal goat serum for 1 h at room temperature, rinsed four times with PBS, and incubated with secondary antibodies produced in goat (diluted 1:400 in 5% normal goat serum) for 45 min at room temperature in the dark. Slides were mounted on glass slides using Vectashield (Vector Laboratories) and imaged on a spinning disk confocal system (Perkin Elmer) or a Zeiss Laser Scanning Microscope (LSM) 710. Images were processed using ImageJ and Adobe PhotoshopCS4. For quantification of lysosome diameter, ImageJ software was used to draw lines across 130–290 lysosomes from 3–5 fields containing 4–6 cells per field. Line pixel values were converted into μ m based on objective magnification and camera pixel size. Presented data are representative of 3–5 replicate experiments. DQ-BSA experiments were performed as previously described³³. Briefly, cells were incubated with DQ-BSA for 30 min and subsequently chased for 1 h in DQ-BSA-free media. Where indicated, treatment with BafA1 was for 1 h at 100 nM. The number of DQ-BSA spots co-localizing with LAMP2-positive lysosomes was counted. Measurement of autolysosome maturation was performed in cells transfected with the mCherry-GFP-LC3 reporter. Cells were transfected with siCTRL or siMITF/siTFE3 and 48 h later transfected with the mCherry-GFP-LC3. After 24 h, cells were fixed in 4% PFA and quantification of total number of autolysosomes (mCherry⁺/GFP[−] spots) compared to total spots per cell ($N > 1,000$ total spots from $N = 10$ cells per condition) for at least three independent experiments was counted.

Histology and immunostaining. Tissue samples were fixed overnight in 4% buffered formaldehyde, and then embedded in paraffin and sectioned (5 mm thickness) by the DF/HCC Research Pathology Core. Haematoxylin and eosin staining was performed using standard methods. For immunohistochemistry, unstained slides were baked at 55 °C overnight, deparaffinized in xylene (two treatments, 6 min each), rehydrated sequentially in ethanol (5 min in 100%, 3 min in 95%, 3 min in 75%, and 3 min in 40%), and washed for 5 min in 0.3% Triton X-100/PBS (PBST) and 3 min in water. For antigen unmasking, specimens were cooked in a 2100 Antigen Retriever (Aptum Biologics) in 1× Antigen Unmasking Solution, Citric Acid Based (H-3300, Vector Laboratories), rinsed three times with PBST, incubated for 10 min with 1% H₂O₂ at room temperature to block endogenous peroxidase activity, washed three times with PBST, and blocked with 5% goat serum in PBST for 1 h. Primary antibodies were diluted in blocking solution as follows: anti-TFE3 (Cell Marque, 354R-14) 1:100; anti-CTSS (Sigma, HPA002988) 1:200; anti-ATP6V1H (Sigma, HPA023421) 1:100; anti-ULK2 (Sigma, HPA009027) 1:100, and incubated with the tissue sections at 4 °C overnight. The evaluation of MITF and TFEB protein levels in patient PDA tissue could not be evaluated due to lack of antibodies suitable for immunohistochemistry (IHC). Commercially available MITF antibodies for IHC are specific to the melanoma (MITF-M) isoform or cross-react with other MIT/TFE family members. Specimens were then washed three times for 3 min each in PBST and incubated with biotinylated secondary antibody (Vector Laboratories) in blocking solution for 1 h at room temperature. Specimens were then washed three times in PBST and treated with ABC reagent (Vectastain ABC kit #PK-6100) for 30 min, followed by three washes for 3 min each. Finally, slides were stained for peroxidase for 3 min with the DAB (diaminebenzidine) substrate kit (SK-4100, Vector Laboratories), washed with water and counterstained with haematoxylin. Immunostaining was assessed by semi-quantitative analysis of tissue microarray, quantifying the intensity of stained tumour cells scored on a scale from 0 (no staining) to 3 (strongest intensity). Stained slides were photographed with an OlympusDP72 microscope.

For human samples, PDA samples used for IHC and generation of early passage PDA cultures were obtained under IRB-approved protocols 02-240 and 2007P001918. The samples for IHC were from patients resected at MGH between 2001 and 2012. Pathological diagnosis of PDA was confirmed by a gastrointestinal cancer pathologist at MGH (V.D.). Early passage PDA cultures were derived from ascites specimens. Patients were verified to have PDA from biopsies.

Electron microscopy. For *in vitro* studies, tissue culture specimens in confluent 6-well or 10-cm plates were fixed directly with gentle rocking for 15 min at room temperature with a glutaraldehyde fixative (2.5% glutaraldehyde, 2% paraformaldehyde, 0.025% calcium chloride, in a 0.1 M sodium cacodylate buffer, pH 7.4), gently scraped and pelleted at 500g, and washed twice with cacodylate buffer. To make a cell block, the cells were centrifuged and resuspended in warm (60 °C) 2% agar in a warm water bath to keep the agar fluid. The cells were then centrifuged again and the agar allowed to harden in an ice water bath. The tissue-containing tip of the centrifuge tube was cut off, resulting in an agar block with the cells embedded within it. This agar block was subsequently routinely processed in a Leica Lynx automatic tissue processor. Briefly, the tissue was post-fixed in osmium tetroxide, stained *en bloc* with uranyl acetate, dehydrated in graded ethanol solutions, infiltrated with propylene oxide/Epon mixtures, embedded in pure Epon, and polymerized overnight at 60 °C. One-micrometre sections were cut, stained with toluidine blue, and examined by light microscopy. Representative areas were chosen for electron microscopic study and the Epon blocks trimmed accordingly. Thin sections were cut with an LKB 8801 ultramicrotome and diamond knife, stained with lead citrate, and examined in a FEI Morgagni transmission electron microscope. Images were captured with an AMT digital CCD camera.

For analysis of human tissue, fresh biopsy tissue placed into Karnovsky's KII Solution (2.5% glutaraldehyde, 2.0% paraformaldehyde, 0.025% calcium chloride, in a 0.1 M sodium cacodylate buffer, pH 7.4), fixed overnight at 4 °C, and stored in cold buffer. Subsequently, they were post-fixed in osmium tetroxide, stained *en bloc* with uranyl acetate, dehydrated in graded ethanol solutions, infiltrated with propylene oxide/Epon mixtures, flat embedded in pure Epon, and polymerized overnight at 60 °C. One-micrometre sections were cut, stained with toluidine blue, and examined by light microscopy. Representative areas were chosen for electron microscopic study and the Epon blocks were trimmed accordingly. Thin sections were cut with an LKB 8801 ultramicrotome and diamond knife, stained with Sato's lead, and examined in a FEI Morgagni transmission electron microscope. Images were captured with an Advanced Microscopy Techniques 2K digital CCD camera.

Measurement of lysosomal pH. Determination of lysosomal pH was carried out on intact cells by ratiometric imaging of the pH-sensitive dye Oregon Green 514, as described^{34,35}. 8988T cells were transfected with control siRNA or TFE3 siRNA. One day before the experiment, 80,000 cells were plated in black 96-multiwell plates in low serum and low glucose media. On the day of the experiment,

cells were fed 30 µg ml⁻¹ 70kDa dextran conjugated to Oregon Green 514 (Dx-OG514, Invitrogen) for 6 h in full media. Cells were rinsed three times to remove excess dye, and chased for 2 h in media free of Dx-OG514, which results in the selective accumulation of Dx-OG514 in lysosomes. Dx-OG514 fluorescence was collected at 530 nm upon excitation at 440 nm and 490 nm in a Spectramax microplate reader (Molecular Devices). The 490/440 fluorescence emission ratios were interpolated to a calibration curve. The calibration curve was built by bathing cells in a K⁺ isotonic solution (145 mM KCl, 10 mM glucose, 1 mM MgCl₂, and 20 mM of either HEPES, MES, or acetate) buffered to pH ranging from 3.5 to 7.0 and containing 10 µg ml⁻¹ Nigericin. The 490/440 ratios were plotted as a function of pH, and fitted to a Boltzmann sigmoid. All measurements were performed in quadruplicate. Presented data are representative of 3 replicate experiments.

Quantitative RT-PCR. Total cellular RNA was extracted using RNeasy Mini Kit (Qiagen) and reverse transcription was performed from 2 µg of total RNA using the QuantiTect Reverse Transcription Kit (Qiagen). Quantitative RT-PCR was performed with FastStart Universal SYBR Green (Roche) in a MX3005P continuous fluorescence detector (Stratagene) or using a Lightcycler 480 (Roche). PCR reactions were performed in triplicate and the relative amount of cDNA was calculated by the comparative CT method using the 18S ribosomal RNA sequences as a control. Primer sequences are available upon request. MITF isoform-specific PCR was conducted as previously described³⁶.

ChIP. ChIP experiments were conducted using the Chromatin Immunoprecipitation assay kit (Millipore) according to the manufacturer's instructions. Briefly, 10⁶ cells expressing Flag-tagged MITF or TFE3 were fixed with 1% formaldehyde for 10 min (for Flag-MITF experiments, cells were fixed 48 h after doxycycline treatment). Cells were washed twice with ice-cold PBS and lysed in ice-cold SDS lysis buffer containing protease inhibitors (Roche). Chromatin extraction and DNA sonication were conducted according to the manufacturer's protocol. DNA was recovered from immune complexes on M2-Flag affinity gel. The immuno-precipitated DNA was recovered and analysed by real-time PCR. ChIP primers have been previously described³⁷.

Xenograft studies. Mice were housed in pathogen-free animal facilities. All experiments were conducted under protocol 2005N000148 approved by the Subcommittee on Research Animal Care at Massachusetts General Hospital. For subcutaneous xenografts, Panc1, 8988T or PL18 cells were infected with lentiviral shRNAs targeting TFE3 (*n* = 2), MITF (*n* = 2) and GFP (control hairpin, *n* = 1) and subjected to a puromycin selection (2 µg ml⁻¹) *in vitro*. 1.5–2.5 × 10⁶ cells, suspended in a 1:1 mixture of HBSS:matrigel, were injected subcutaneously into the lower flank of 3–4 female NOD.CB17-Prkdcscid/J mice per group (6–10 weeks of age) (purchase from Jackson Laboratories, strain #001303). Tumour length and width were measured weekly and the volume was calculated according to the formula (length × width²)/2. All xenograft experiments with human PDA lines were approved by the HMS Institutional Animal Care and Use Committee (IACUC) under protocol number 04-605. For orthotopic mouse studies, MITF or GFP was stably expressed in mouse PanIN primary cells³⁸. Alternatively, a doxycycline-inducible expression construct for MITF was also stably expressed in mouse PanIN primary cells. 1 × 10⁶ cells were orthotopically injected in a 1:1 mixture of matrigel:PBS into the pancreas of SCID mice (*n* = 5 per group). Mice were euthanized 6 weeks after implantation, pancreas were collected and submitted for histological examination. No mice were excluded from the analysis. Experiments were designed to detect a 50% change in tumour size with 80% power and a type I error of 5% using the *t*-test. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Image analysis and statistics. Image analysis, including densitometry and measurement of fluorescence intensity, was conducted using Image J software (NIH). Statistical analyses of results are expressed as mean ± standard deviation, unless otherwise specified. Significance was analysed using two-tailed Student's *t*-test. A *P* value of less than 0.05 was considered statistically significant.

Gene expression profiling and GSEA. To build a comprehensive autophagy-lysosome gene signature (gene set), we combined published lysosome proteomics^{39,40} and autophagy interactome data sets⁴¹ together with known lysosomal disease associated genes⁴² (see Supplementary Table 1). Data sets used for the meta-analysis in Fig. 1c, f and Extended Data Fig. 1b, c are accessible from the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/gds/>), including GSE16515, GSE28735 and GSE15471, from The Cancer Genome Atlas (TCGA; <http://cancergenome.nih.gov/>), from the Cancer Cell Line Encyclopedia (CCLE; <http://www.broadinstitute.org/software/cprg/?q=node/11>). For the GEO and CCLE data sets, raw expression values in the form of CEL files were collected and then processed using RMA in the R bioconductor package. For TCGA data, expression data sets were created by combining RNASeqV2 Level3 normalized gene result files for individual samples and producing tables with genes in rows and samples in columns. Data for the 8988T cells of Fig. 1g, h was processed using

a standard RNA-seq pipeline that used Trimmomatic to clip and trim the reads, used tophat2 to align the reads to hg19, and used cuffdiff to calculate differential expression. GSEA (<http://www.broadinstitute.org/gsea/index.jsp>) of the expression data was used to assess enrichment of the autophagy-lysosome gene signature. Depending upon the data set, there were several different methods used to rank genes for GSEA. In the PDA samples for Fig. 1f and Extended Data Fig. 1i, j genes were ranked according to Pearson correlation with a meta-gene formed by the mean expression of MITF, TFE3 and TFEB, and *P* values were obtained by permuting the phenotype (2,500 permutations). In the 8988T cells of Fig. 1g, h, a pairwise GSEA was performed by creating ranked lists of genes using the log₂ ratio of siTFE3 to siCTRL, and *P* values were obtained by permuting the gene set (1,000 permutations). In the paired tumour/normal PDA data sets of Extended Data Fig. 1b, a paired *t*-test between matched tumour-normal samples was used to rank genes and the ranked list was used in GSEA with *P* values from permuting the gene set (2,500 permutations). For Extended Data Fig. 1k, Database for Annotation, Visualization and Integrated Discovery (DAVID) analysis of PDA cell lines in CCLE was performed as described⁴³ by using Pearson correlation to select the 100 genes most correlated with TFE3 and MITF, and checking for enriched pathways. **Quantitative proteomics on TFE3-associated proteins.** Enriched proteins were subjected to multiplexed quantitative proteomics analysis using tandem-mass tag (TMT) reagents on an Orbitrap Fusion mass spectrometer (Thermo Scientific). Disulfide bonds were reduced with dithiothreitol (DTT) and free thiols alkylated with iodoacetamide as described previously⁴⁴. Proteins were then precipitated with trichloroacetic acid, resuspended in 50 mM HEPES (pH 8.5) and 1 M urea and digested first with endoproteinase Lys-C (Wako) for 17 h at room temperature and then with sequencing-grade trypsin (Promega) for 6 h at 37 °C. Peptides were desalted over Sep-Pak C₁₈ solid-phase extraction (SPE) cartridges (Waters), the peptide concentration was determined using a BCA assay (Thermo Scientific) and a maximum of 50 µg of peptides were labelled with one out of the available TMT-10plex reagents (Thermo Scientific)⁴⁵. To achieve this, peptides were dried and resuspended in 50 µl of 200 mM HEPES (pH 8.5) and 30% acetonitrile (ACN) and 10 µg of the TMT in reagent in 5 µl of anhydrous ACN was added to the solution, which was incubated at room temperature for 1 h. The reaction was then quenched by adding 6 µl of 5% (w/v) hydroxylamine in 200 mM HEPES (pH 8.5) and incubation for 15 min at room temperature. The solutions were acidified by adding 50 µl of 1% trifluoroacetic acid (TFA) and samples that were quantified in one mass spectrometry measurement were pooled before being subjected to desalting over a Sep-Pak C₁₈ (SPE) cartridge. The peptide mixtures were then dried and resuspended in 8 µl, three of which were analysed by microcapillary liquid chromatography tandem mass spectrometry on an Orbitrap Fusion mass spectrometer and using a recently introduced multistage (MS3) method to provide highly accurate quantification^{44,45}.

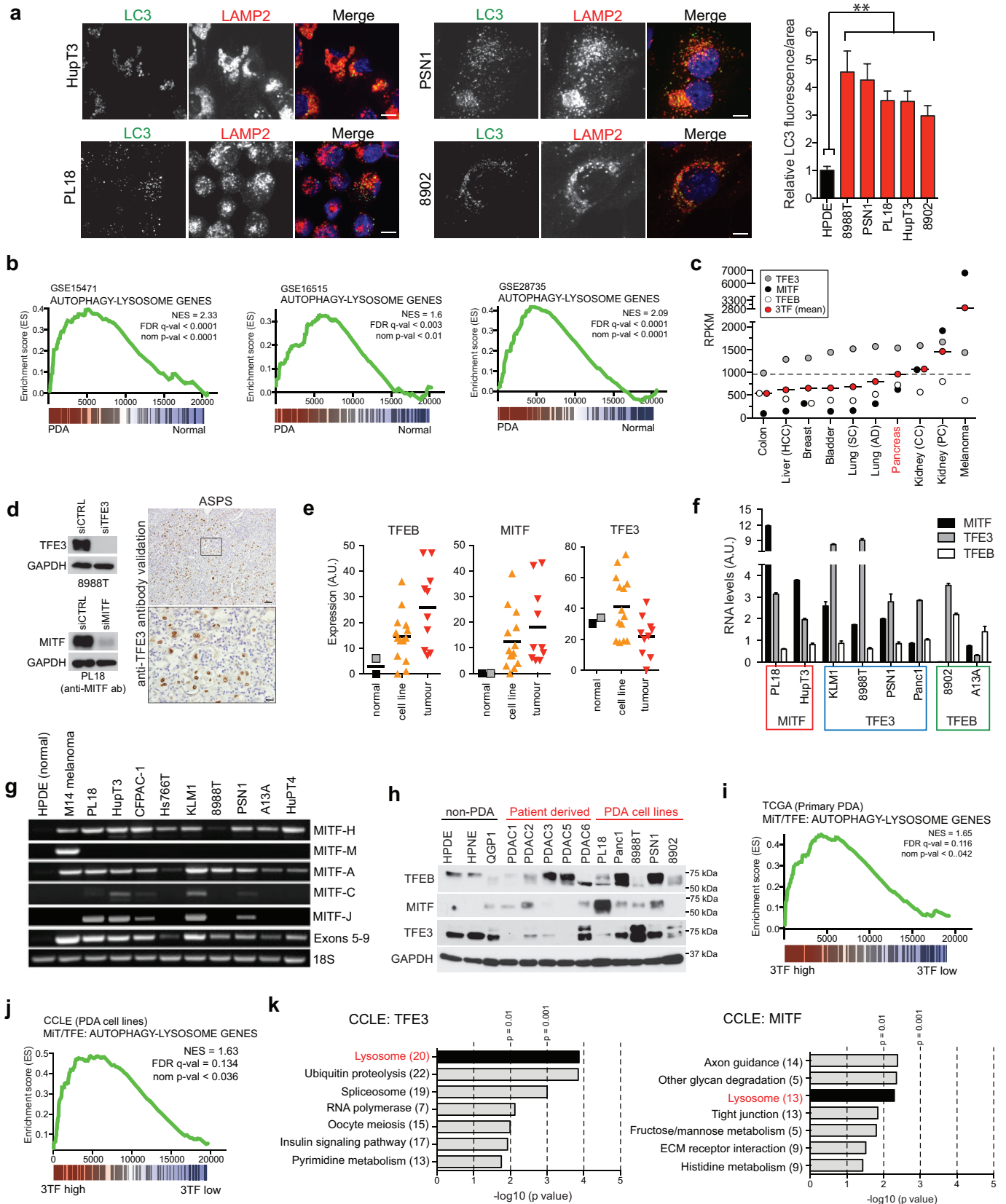
The mass spectrometer was equipped with an EASY-nLC 1000 integrated autosampler and HPLC pump system. Peptides were separated over a 100 µm inner diameter microcapillary column in-house packed with first 0.5 cm of Magic C4 resin (5 µm, 100 Å, Michrom Bioresources), then with 0.5 cm of Maccel C₁₈ resin (3 µm, 200 Å, Nest Group) and 29 cm of GP-C18 resin (1.8 µm, 120 Å, Sepax Technologies). Peptides were eluted applying a gradient of 8–27% ACN in 0.125% formic acid over 165 min at a flow rate of 300 nl min⁻¹. To identify and quantify the TMT-labelled peptides we applied a synchronous precursor selection MS3 method^{44–46} in a data-dependent mode. The scan sequence was started with the acquisition of a full MS or MS1 one spectrum acquired in the Orbitrap (*m/z* range, 500–1,200; resolution, 60,000; AGC target, 5 × 10⁵; maximum injection time, 100 ms), and the ten most intense peptide ions from detected in the full MS spectrum were then subjected to MS2 and MS3 analysis, while the acquisition time was optimized in an automated fashion (top speed, 5 s). MS2 scans were done in the linear ion trap using the following settings: quadrupole isolation at an isolation width of 0.5 Th; fragmentation method, CID; AGC target, 1 × 10⁴; maximum injection time, 35 ms; normalized collision energy, 30%. Using synchronous precursor selection the ten most abundant fragment ions were selected for the MS3 experiment following each MS2 scan. The fragment ions were further fragmented using the HCD fragmentation (normalized collision energy, 50%) and the MS3 spectrum was acquired in the Orbitrap (resolution, 60,000; AGC target, 5 × 10⁴; maximum injection time, 250 ms).

Data analysis was performed on an in-house generated SEQUEST-based⁴⁷ software platform. RAW files were converted into the mzXML format using a modified version of ReAdW.exe. MS2 spectra were searched against a protein sequence database containing all protein sequences in the human UniProt database (downloaded 4 February 2014) as well as that of known contaminants such as porcine trypsin. This target component of the database was followed by a decoy component containing the same protein sequences but

in flipped (or reversed) order⁴⁸. MS2 spectra were matched against peptide sequences with both termini consistent with trypsin specificity and allowing two missed trypsin cleavages. The precursor ion *m/z* tolerance was set to 50 p.p.m., TMT tags on the N terminus and on lysine residues (229.162932 Da) as well as carbamidomethylation (57.021464 Da) on cysteine residues were set as static modification, and oxidation (15.994915 Da) of methionines as variable modification. Using the target-decoy database search strategy⁴⁸ a spectra assignment false discovery rate of less than 1% was achieved through using linear discriminant analysis with a single discriminant score calculated from the following SEQUEST search score and peptide sequence properties: mass deviation, XCorr, dCn, number of missed trypsin cleavages, and peptide length⁴⁹. The probability of a peptide assignment to be correct was calculated using a posterior error histogram and the probabilities for all peptides assigned to a protein were combined to filter the data set for a protein FDR of less than 1%. Peptides with sequences that were contained in more than one protein sequence from the UniProt database were assigned to the protein with most matching peptides⁴⁹.

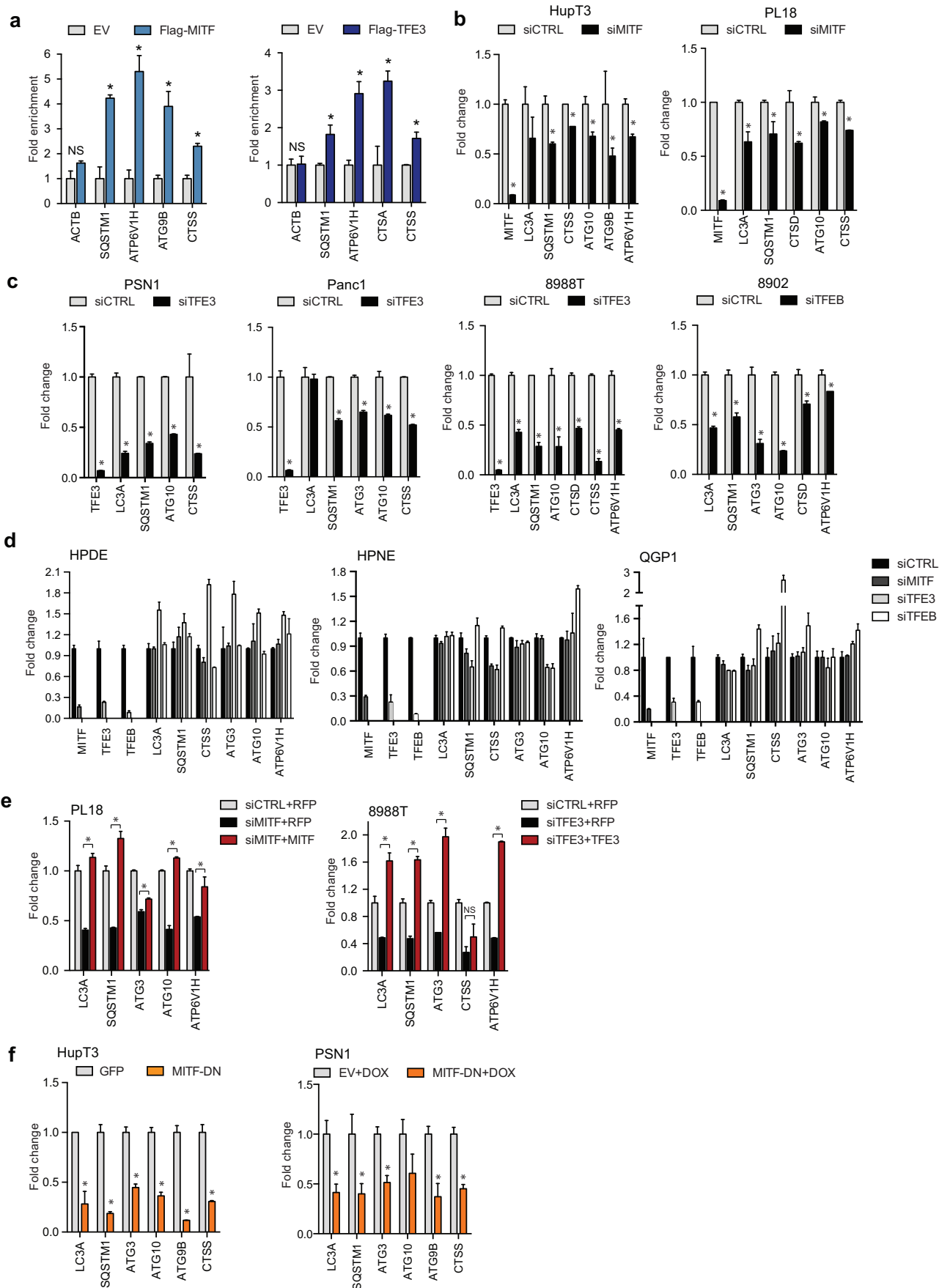
TMT reporter ion intensities were extracted as that of the most intense ion within a 0.03 Th window around the predicted reporter ion intensities in the collected MS3 spectra. Only MS3 with an average signal-to-noise value of larger than 28 per reporter ion as well as with an isolation specificity⁴⁴ of larger than 0.75 were considered for quantification. Reporter ions from all peptides assigned to a protein were summed to define the protein intensity. A two-step normalization of the protein TMT intensities was performed by first normalizing the protein intensities over all acquired TMT channels for each protein based to the median average protein intensity calculated for all proteins. To correct for slight mixing errors of the peptide mixture from each sample a median of the normalized intensities was calculated from all protein intensities in each TMT channel and the protein intensities were normalized to the median value of these median intensities.

29. Tzatsos, A. *et al.* KDM2B promotes pancreatic cancer via Polycomb-dependent and -independent transcriptional programs. *J. Clin. Invest.* **123**, 727–739 (2013).
30. Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
31. Gameiro, P. A. *et al.* *In vivo* HIF-mediated reductive carboxylation is regulated by citrate levels and sensitizes VHL-deficient cells to glutamine deprivation. *Cell Metab.* **17**, 372–385 (2013).
32. Metallo, C. M. *et al.* Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* **481**, 380–384 (2012).
33. Commisso, C. *et al.* Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497**, 633–637 (2013).
34. Steinberg, B. E. *et al.* A cation counterflux supports lysosomal acidification. *J. Cell Biol.* **189**, 1171–1186 (2010).
35. Zoncu, R. *et al.* mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H⁺-ATPase. *Science* **334**, 678–683 (2011).
36. Hershey, C. L. & Fisher, D. E. Genomic analysis of the *Microphthalmia* locus and identification of the MITF-J/Mitf-J isoform. *Gene* **347**, 73–82 (2005).
37. Settembre, C. *et al.* TFEB links autophagy to lysosomal biogenesis. *Science* **332**, 1429–1433 (2011).
38. Corcoran, R. B. *et al.* STAT3 plays a critical role in KRAS-induced pancreatic tumorigenesis. *Cancer Res.* **71**, 5020–5029 (2011).
39. Chapel, A. *et al.* An extended proteome map of the lysosomal membrane reveals novel potential transporters. *Mol. Cell. Proteomics* **12**, 1572–1588 (2013).
40. Schröder, B. *et al.* Integral and associated lysosomal membrane proteins. *Traffic* **8**, 1676–1686 (2007).
41. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466**, 68–76 (2010).
42. Ballabio, A. & Gieselmann, V. Lysosomal disorders: from storage to cellular damage. *Biochim. Biophys. Acta* **1793**, 684–696 (2009).
43. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* **4**, 44–57 (2009).
44. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8**, 937–940 (2011).
45. McAlister, G. C. *et al.* Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478 (2012).
46. Weekes, M. P. *et al.* Quantitative temporal viromics: an approach to investigate host-pathogen interaction. *Cell* **157**, 1460–1472 (2014).
47. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
48. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).
49. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
50. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).



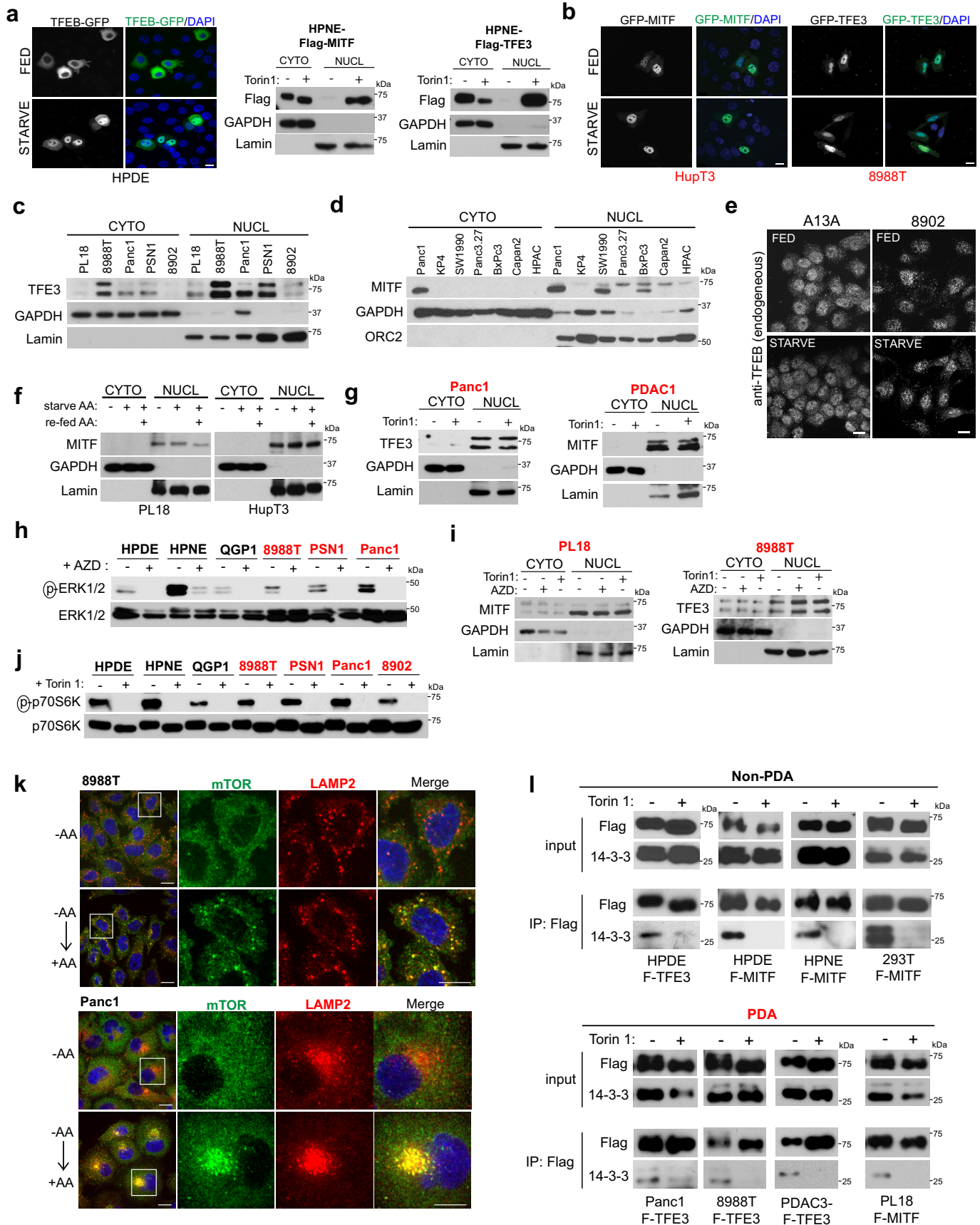
Extended Data Figure 1 | Tumour-specific expression and constitutive activation of MiT/TFE proteins in PDA. **a**, Immunofluorescence staining of autophagosomes (LC3) and lysosomes (LAMP2) showing extensive overlap of these organelles and increased LC3 immunofluorescence in PDA cell lines (graph, right; error bars indicate mean \pm s.d. for $N = 3$ independent experiments with at least 130 cells scored. $^{**}P < 0.001$. Significance was analysed using two-tailed Student's *t*-test). Scale bar, 11 μ m. **b**, GSEA of different human PDA data sets for enrichment of the autophagy-lysosome gene signature in tumour versus normal tissue (Gene Expression Omnibus (GEO) accession numbers indicated). **c**, Mean expression (reads per kilobase per million (RPKM)) of TFE3, MITF and TFEB individually or cumulatively as a meta-gene formed by the mean of the three transcription factors (3TF) in primary tumour specimens from the indicated malignancies (The Cancer Genome Atlas (TCGA) data set). **d**, TFE3 and MITF antibody validation by western blotting in cells treated with the indicated siRNA. TFE3 antibody immunohistochemistry (IHC) validation using alveolar soft part sarcoma (ASPS) tissue. This antibody (MRQ-37, Cell Marque) is used as a clinical diagnostic for ASPS. Scale bars: 100 μ m (top); 20 μ m (bottom). **e**, Gene expression analysis showing upregulation of MiT/TFE genes in subsets of PDA relative to normal pancreatic ductal tissue. SAGE data³⁰ from normal

microdissected pancreatic ductal cells (normal microdissected control, black box; HPDE, grey box), cultured PDA cells (cell line), and PDA xenografts and primary tumour tissues (tumour). A.U., arbitrary units. **f**, Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis of MiT/TFE expression levels in a panel of human PDA cell lines. Note that PL18 and HupT3 preferentially express high levels of MITF, while 8988T, PSN1 and Panc1 express TFE3 at higher levels. The highest levels of TFEB were detected in 8902 and A13A cells. **g**, RT-PCR analysis reveals that PDA cells express distinct MITF isoforms. Note the complete absence of all MITF isoforms in normal HPDE cells. PDA cells lack the melanoma-specific M isoform (detected in M14 melanoma cells, lane 2). **h**, MITF, TFE3 and TFEB protein levels in a panel of non-PDA cell lines, patient-derived PDA cultures and PDA cell lines. **i**, GSEA analysis showing correlation between expression of MiT/TFE factors and autophagy-lysosome gene set in primary human PDA specimens (TCGA data set). **j**, GSEA analysis showing correlation between cumulative expression of MiT/TFE factors (see Methods) and the autophagy-lysosome gene set in human PDA cell lines (Cancer Cell Line Encyclopedia (CCLE) data set). **k**, Database for Annotation, Visualization and Integrated Discovery (DAVID) analysis of gene sets correlating with increasing expression of TFE3 (left) or MITF (right) in human PDA cell lines (CCLE data set).



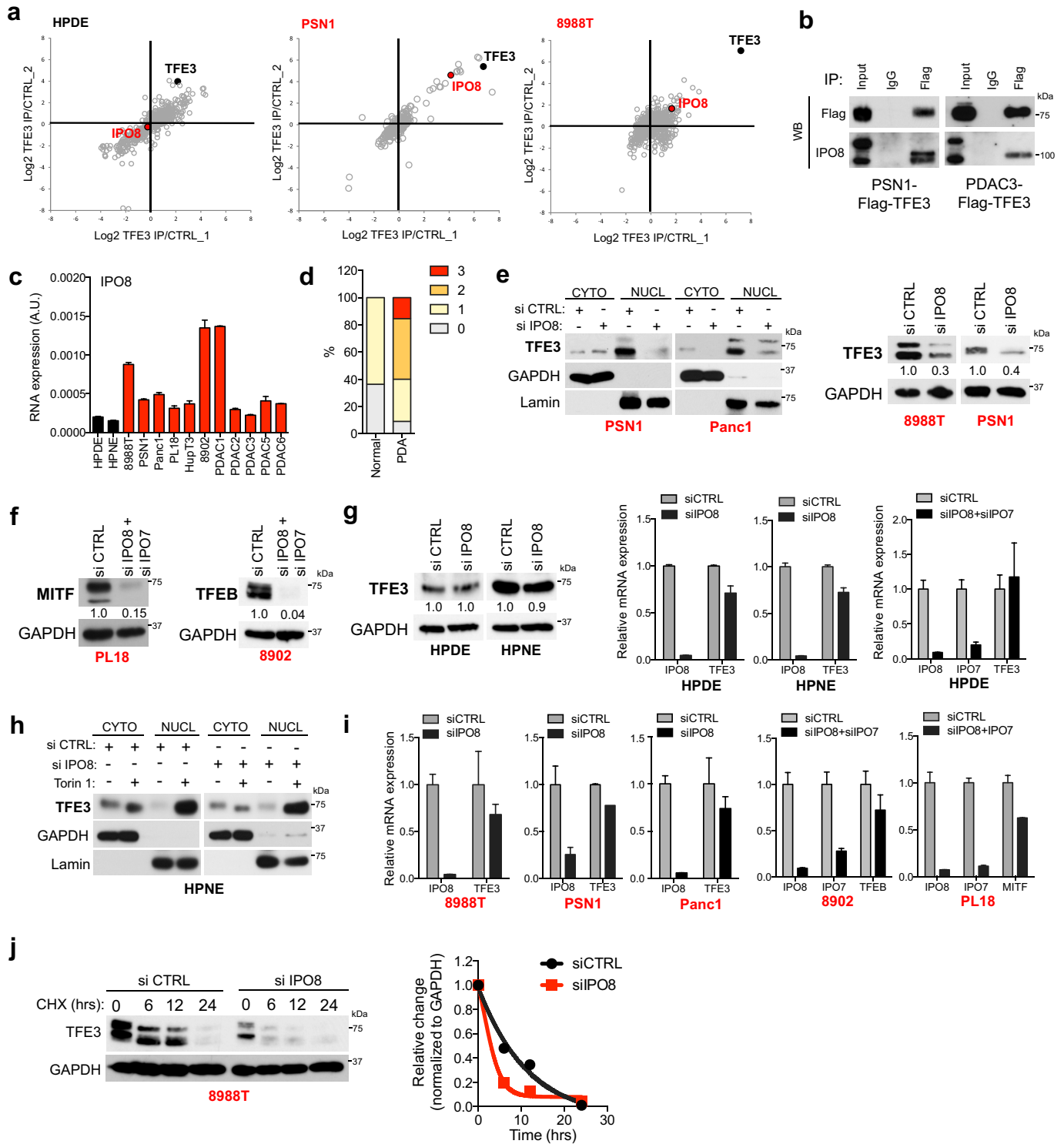
Extended Data Figure 2 | MiT/TFE-dependent regulation of autophagy-lysosome gene expression in PDA cell lines. **a**, ChIP analysis of Flag-MITF (left) and Flag-TFE3 (right) binding to autophagy-lysosome genes in 8902 and Panc1 cells, respectively. Histograms show the amount of immunoprecipitated DNA detected by qPCR normalized to input and plotted as relative enrichment over mock control. Error bars indicate mean \pm s.e.m. for $N = 3$ independent experiments. $*P < 0.05$. EV, empty vector. **b**, siRNA-mediated knockdown of MITF in HupT3 and PL18 cells causes a decrease in autophagy-lysosome gene expression, assayed 48 h after siRNA transfection. $*P < 0.05$. **c**, Knockdown of TFE3 in PSN1, Panc1 and 8988T cells, or of TFEB in 8902 cells, causes a decrease in autophagy-lysosome gene expression. $*P < 0.05$. **d**, HPDE, HPNE and QGP1 control cells show minimal changes in autophagy-lysosome gene expression upon knockdown of the MiT/TFE genes.

e, Decreased autophagy-lysosome gene expression after MITF knockdown (left; PL18 cells; $*P < 0.05$) or TFE3 knockdown (right; 8988T cells; $*P < 0.01$) is rescued by transient ectopic expression of MITF or TFE3. Cells were transfected with expression constructs for MITF or TFE3 24 h post-siRNA transfection. After 48 h, gene expression was assayed. **f**, Expression of dominant-negative MITF (MITF-DN) in HupT3 cells causes a decrease in autophagy-lysosome gene expression compared with control cells (left; $*P < 0.02$). Similar results are seen in PSN1 cells expressing doxycycline (Dox)-inducible MITF-DN upon addition of $1 \mu\text{g ml}^{-1}$ of Dox for 48 h (right; $*P < 0.05$). For all graphs error bars indicate mean \pm s.d. for $N = 3$ independent experiments. Significance was analysed using two-tailed Student's *t*-test.



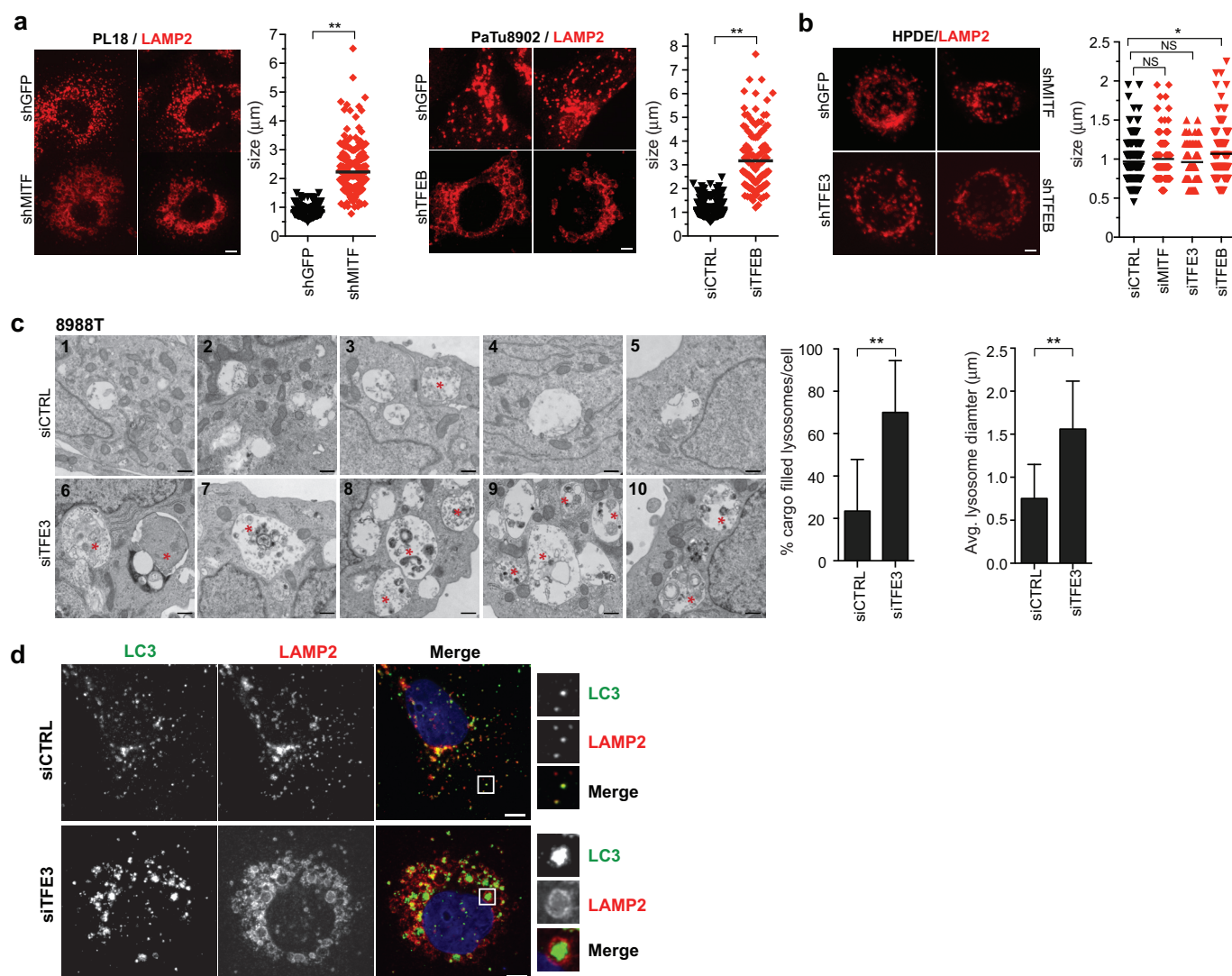
Extended Data Figure 3 | MiT/TFE transcription factors escape mTOR-mediated cytoplasmic retention in PDA. **a**, Subcellular localization of ectopically expressed TFE3-GFP in HPDE cells under full nutrient and starvation conditions (3 h HBSS) (left) and Torin-1-dependent nuclear localization of ectopically expressed Flag-MITF (left) or Flag-TFE3 (right) in HPNE cells (right). **b**, Subcellular localization of ectopically expressed GFP-MITF (left) or GFP-TFE3 (right) in HupT3 and 8988T cells, respectively, under full nutrient and starvation conditions (3 h HBSS). **c**, **d**, Subcellular fractionation studies showing that endogenous TFE3 (**c**) and MITF (**d**) are constitutively nuclear localized in PDA cell lines. **e**, Immunofluorescence staining of endogenous TFE3 in A13A and 8902 PDA cells. Note the predominant nuclear localization under both full nutrient (fed) and starved conditions. **f**, Subcellular fractionation of PL18 and HupT3 PDA cells under full nutrient, amino acid (AA) starved and amino acid re-fed conditions shows constitutive nuclear residence of endogenous MITF regardless of the nutrient status of the cells. Lamin = nuclear fraction; GAPDH = cytoplasmic fraction. **g**, Subcellular fractionation of Panc1 cells and a primary patient-derived culture

(PDAC1) showing constitutive nuclear localization of TFE3 (in Panc1) and MITF (in PDAC1) independent of Torin 1 treatment. **h**, Immunoblot for p-ERK1/2 in the indicated cell lines treated with vehicle or with the MEK inhibitor, AZD6244 (AZD). **i**, Neither AZD6244 nor Torin 1 affect MITF localization (PL18 cells) or TFE3 localization (8988T cells). **j**, Immunoblot showing readily detectable p-p70S6K in the indicated non-PDA and PDA cell lines, and extinction of phosphorylation upon Torin 1 treatment. **k**, Immunofluorescence showing that amino acid re-feeding of starved 8988T and Panc1 PDA cells results in mTOR (green) translocation from a diffuse cytoplasmic distribution to the lysosome (LAMP2; red), as indicated by colocalization with LAMP2. Scale bar, 20 μ m. **l**, The indicated non-PDA (top) and PDA cell lines (bottom) stably expressing Flag-tagged TFE3 (F-TFE3) or MITF (F-MITF) were treated with vehicle or Torin 1. Cells were then lysed, subjected to Flag immunoprecipitation and immunoblotted for Flag and 14-3-3. Note that in all cell lines, 14-3-3 is detected in the anti-Flag immunoprecipitates and binding is lost upon Torin 1 treatment. All data shown are representative of at least $N = 3$ independent experiments.



Extended Data Figure 4 | IPO8 drives increased MiT/TFE nuclear import in PDA cells. **a**, Identification of IPO8 as a PDA-specific binding partner of TFE3. HPDE (left), PSN1 (middle) and 8988T cells (right) stably expressing Flag-TFE3 or control vector were subjected to affinity purification, followed by multiplexed quantitative proteomics analysis using tandem-mass tag (TMT) reagents. The graphs show normalized protein intensities in the Flag-TFE3 and control samples. Note the specific enrichment of IPO8 in the Flag-TFE3-expressing PDA cell lines. In HPDE cells, IPO8 did not score as a significantly enriched interactor (mean \log_2 ratio Flag-TFE3/control = -0.14 ($N = 3$), $P = 0.30$ (paired two-tailed t -test), while in PSN1 and 8988T IPO8 was significantly enriched in TFE3 immunoprecipitates (mean \log_2 ratio Flag-TFE3/control = 4.35 ($N = 3$), $P = 0.015$ (PSN1) and mean \log_2 ratio Flag-TFE3/control = 1.70 ($N = 3$), $P = 0.002$ (8988T)). **b**, Immunoprecipitation of endogenous IPO8 with Flag-TFE3 in a PDA cell line (PSN1) and primary PDA culture (PDAC3). **c**, qRT-PCR showing increased expression of IPO8 in PDAC cell lines and primary patient-derived cultures (red bars) compared to control pancreatic ductal cells (HPDE and HPNE, black bars). **d**, Quantification of IPO8 immunohistochemistry staining intensity (0 = no staining to 3 = high

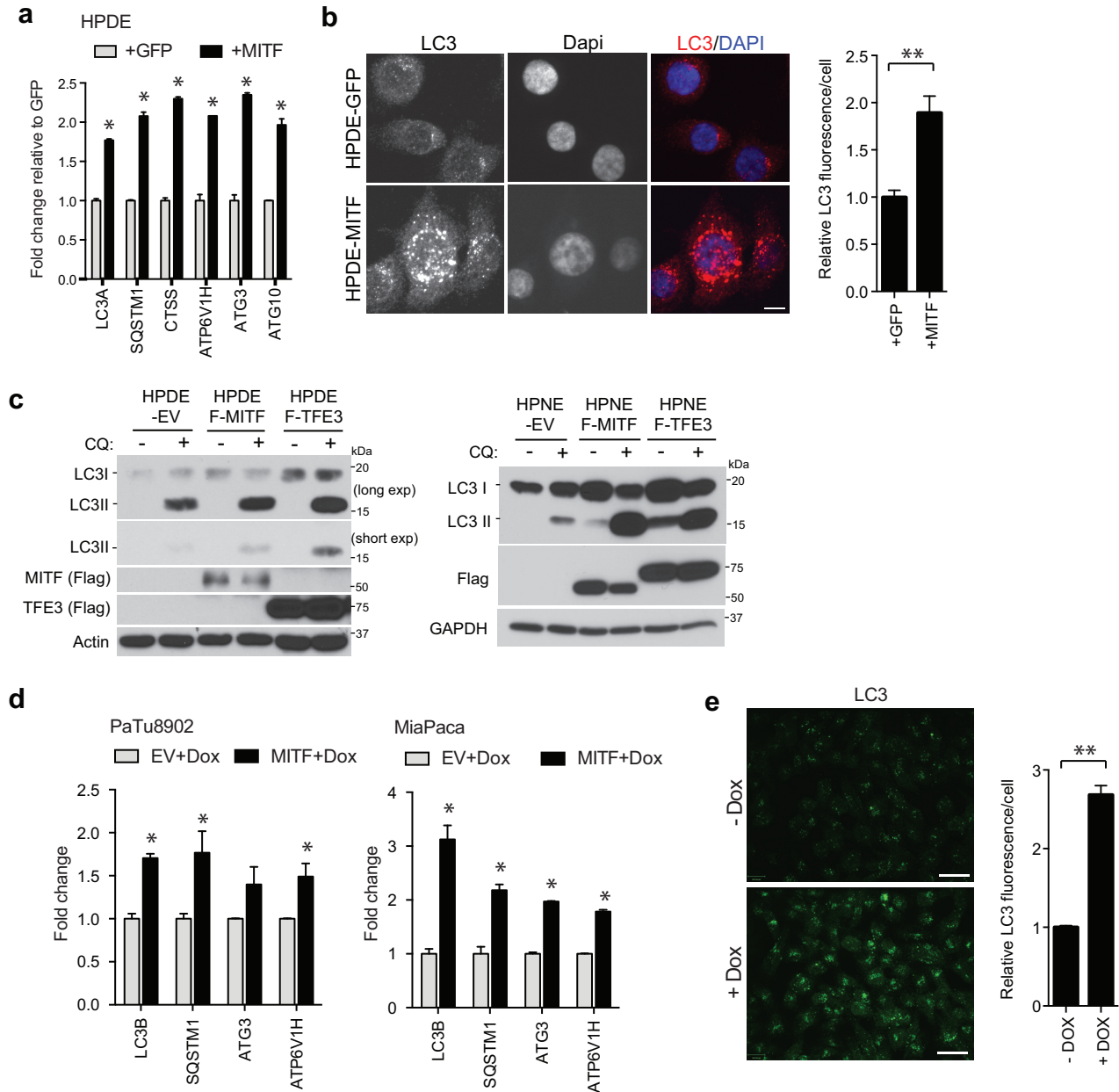
staining) in normal ($N = 11$) and PDA ($N = 110$) patient samples. **e**, Subcellular fractionation and immunoblot analyses of the indicated cell lines transfected with control siRNA (siCTRL) or siIPO8. Note that siIPO8 leads to a marked decrease in nuclear TFE3 in PDA cells (PSN1 and Panc1) (left) and in whole cell lysates (right). **f**, Immunoblot of whole cell lysates showing that IPO8+IPO7 knockdown decreases the levels of MITF and TFEB in PL18 and 8902 cells, respectively. **g**, Knockdown of IPO8 has no effect on total TFE3 protein (left) or mRNA (right) levels in HPDE and HPNE cells. Error bars indicate mean \pm s.d. for $N = 3$ independent experiments. **h**, Torin-1-induced TFE3 nuclear localization in HPNE cells is unaffected by knockdown of IPO8. **i**, qRT-PCR showing that siRNA-mediated knockdown of IPO8, or of both IPO8 and IPO7, effectively reduces target expression without significantly affecting the expression of *TFE3*, *TFEB* or *MITF* mRNA levels. Error bars indicate mean \pm s.d. for $N = 3$ independent experiments. **j**, Immunoblot of 8988T cells transfected with control siRNA (siCTRL) or siIPO8 and treated with cycloheximide (CHX) for the indicated time points shows a decrease in steady-state levels and stability of TFE3 upon loss of IPO8. Data are representative of $N = 3$ independent experiments.



Extended Data Figure 5 | Altered lysosome morphology and function following loss of MiT/TFE proteins. **a**, RNAi-mediated knockdown of MITF in PL18 (right; $N = 259$ siCTRL, $N = 263$ siMITF) or TFEB in PaTu8902 cells (left; $N = 273$ siCTRL, $N = 147$ siTFEB) causes aberrant lysosome morphology and an increase in lysosome diameter as visualized by immunofluorescence staining for LAMP2. $**P < 0.001$; bar, mean. **b**, Lysosome size in HPDE cells is not affected by knockdown of MITF ($N = 156$) and TFE3 ($N = 81$), and is only slightly increased by TFEB knockdown ($N = 198$) relative to siCTRL ($N = 296$). NS, not significant; $*P < 0.05$; bar, mean. Scale bar, $7.5 \mu\text{m}$. **c**, Electron microscopy of 8988T PDA cells transfected with siCTRL (panels 1–5) or siTFE3 (panels 6–10). Note that TFE3 loss causes an accumulation of undigested material shown by asterisks indicating a defect in clearance (graph

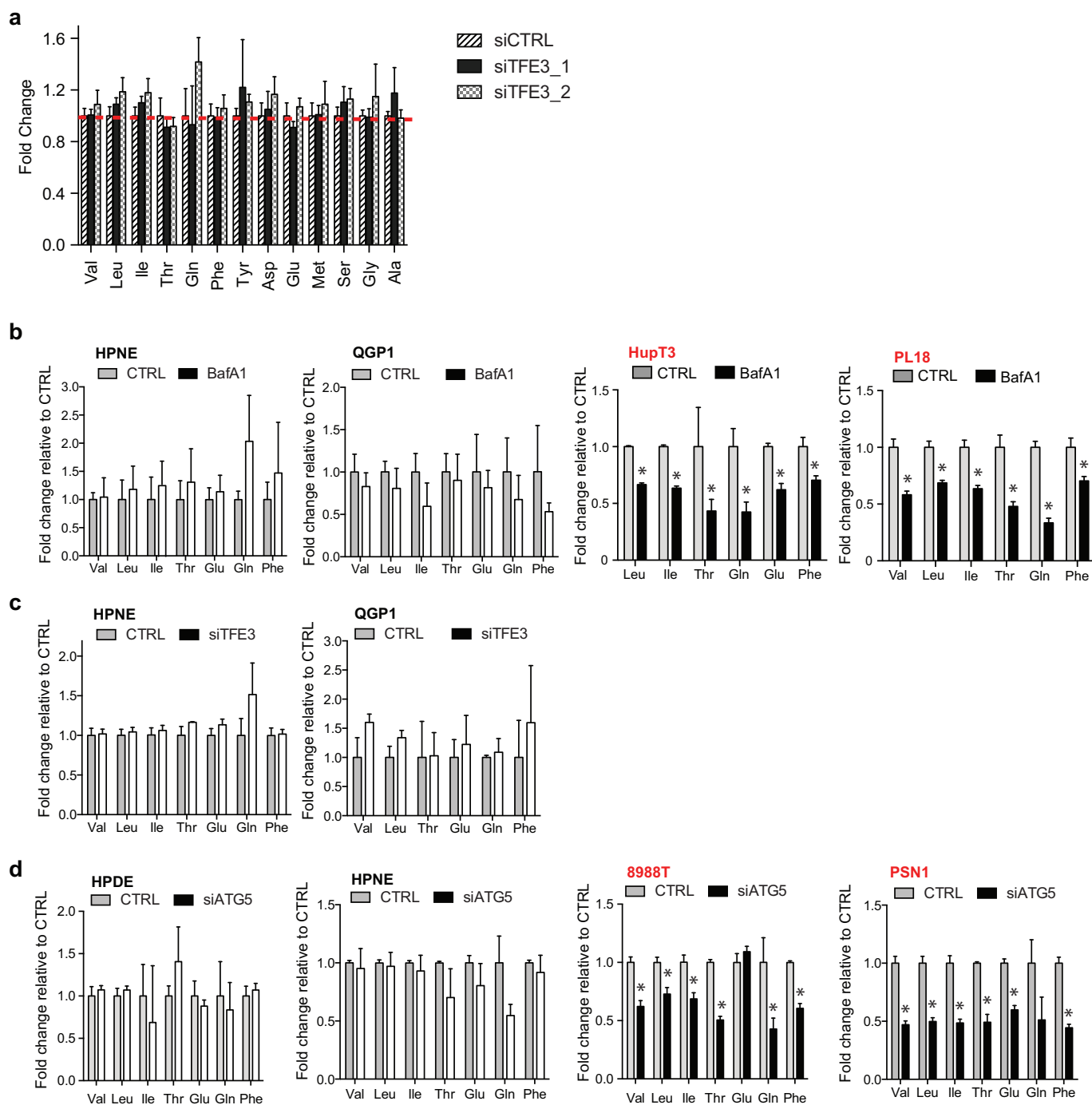
indicates percentage lysosomes filled with cargo; $N = 80$ lysosomes for siCTRL and $N = 153$ lysosomes for siTFE3) and an increase in average lysosome diameter (quantified in graph on the right; $N = 63$ lysosomes in siCTRL and $N = 68$ lysosomes in siTFE3). Scale bar, $1 \mu\text{m}$. $**P < 0.001$.

d, Immunofluorescence staining with LC3 (green) and LAMP2 (red) in 8988T cells after siRNA-mediated knockdown of TFE3, shows similar accumulation of undigested LC3-positive aggregates encapsulated within enlarged LAMP2 positive lysosomes (bottom) compared with control cells (top). Magnifications of the boxed regions are shown (right). Scale bar, $7.5 \mu\text{m}$. Error bars represent mean \pm s.d. Data are representative of at least $N = 3$ independent experiments. Significance was analysed using two-tailed Student's *t*-test.



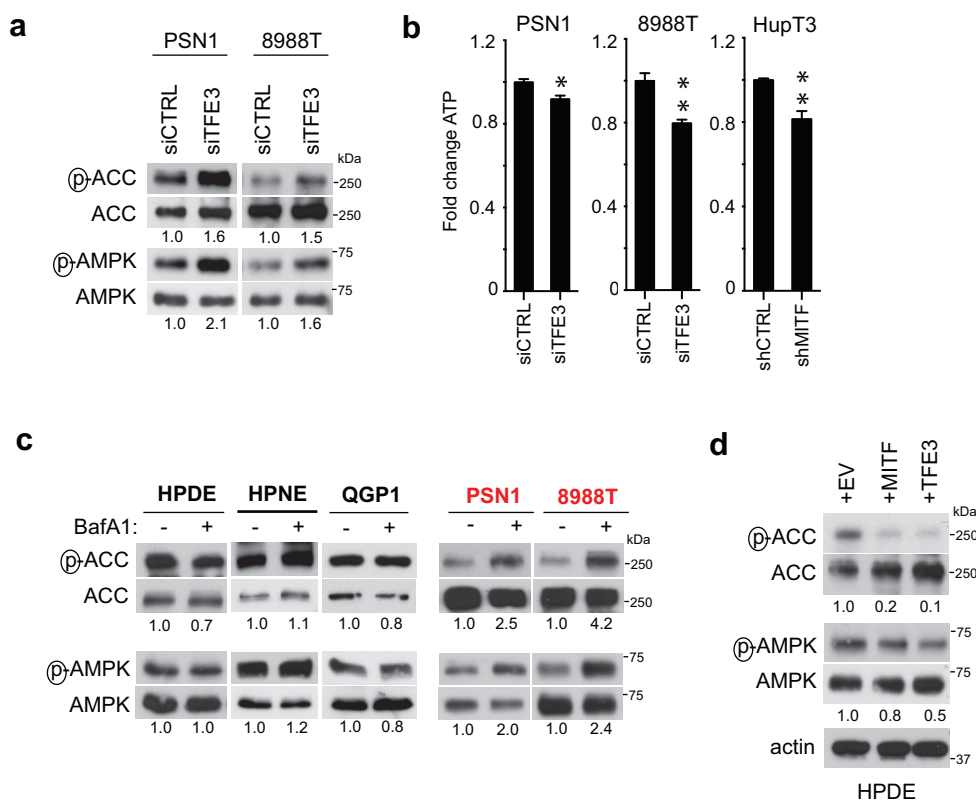
Extended Data Figure 6 | Ectopic Mit/TFE expression in PDA cell lines causes an increase in autophagy-lysosome function. **a, b,** Ectopic expression of MITF induces autophagy-lysosome genes in HPDE cells (**a**), and causes an increased abundance of LC3 puncta (**b**) as measured by immunofluorescence staining of endogenous LC3 (red) and quantified in the graph on the right. $*P < 0.01$, $**P < 0.001$. Scale bar, 7.5 μm . **c,** Ectopic expression of Flag-tagged MITF or TFE3 in HPDE cells (left) or HPNE cells (right) causes an increase in autophagic flux as measured by the increase in LC3-II versus LC3-I, following treatment with 25 μM chloroquine (CQ) for 18 h. MITF and TFE3 protein

expression is represented by immunoblot for Flag. **d,** Expression of MITF in 8902 PDA cells (which lack MITF expression) or in MiaPaca cells (which express low levels of all Mit/TFE members) causes an increase in autophagy-lysosome gene expression. $*P < 0.05$. **e,** Dox-inducible expression of MITF in MiaPaca cells causes an increase in endogenous LC3-positive puncta, as measured in the graph on the right, indicating increased autophagy induction. $N = 68$ cells, -Dox; $N = 100$ cells, +Dox; $**P < 0.001$. Error bars represent mean \pm s.d. Data are representative of at least $N = 3$ independent experiments. Significance was analysed using two-tailed Student's *t*-test. Scale bar, 15 μm .



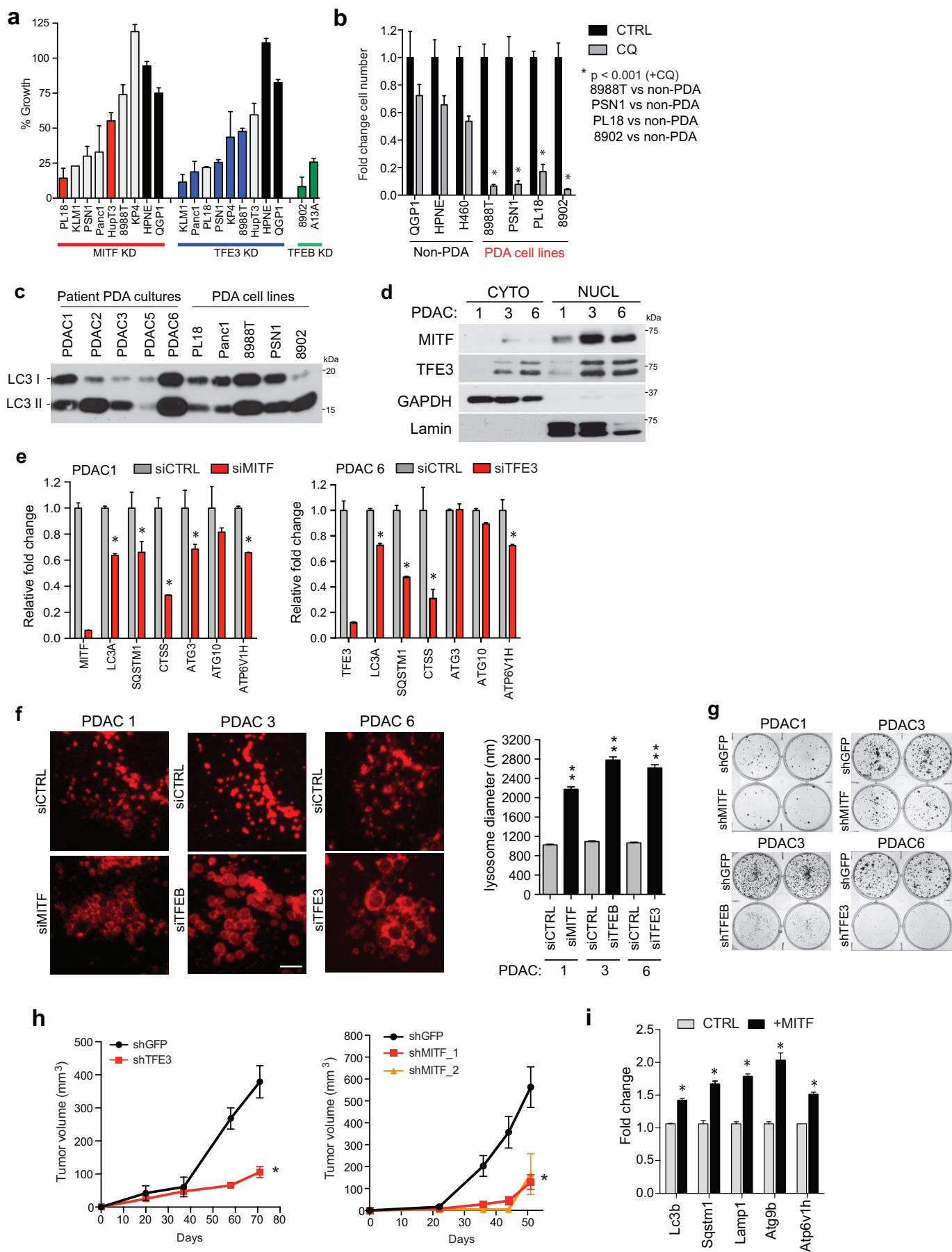
Extended Data Figure 7 | Role of lysosome in maintaining amino acid levels in PDA. **a**, Amino acid uptake was measured as fold change in extracellular amino acids in 8988T cells after transfection with two siRNAs against TFE3 relative to siCTRL. Media was changed 1 h before media samples were harvested for analysis. Data are matched to results presented in Fig. 3b.

b–d, Effect of BafA1 (**b**), siTFE3 (**c**), and siATG5 (**d**) on intracellular amino acid levels in the indicate cell lines. Error bars represent mean \pm s.d. for $N = 3$ independent experiments. $*P < 0.05$. Significance was analysed using two-tailed Student's *t*-test.



Extended Data Figure 8 | MiT/TFE proteins couple amino acid metabolism to energy homeostasis in PDA. **a**, Knockdown of TFE3 in PSN1 and 8988T cells causes an increase in p-ACC (Ser 79) and p-AMPK (Thr 172) levels. **b**, Knockdown of TFE3 (in 8988T and PSN1 cells) or MITF (in HupT3 cells) causes a decrease in cellular ATP levels. $N = 3$ independent experiments; * $P < 0.05$, ** $P < 0.001$. **c**, BafA1 treatment (150 nM) for 18 h induces p-ACC

and p-AMPK in PDA cells but not in HPDE, HPNE or QGP1 cells. **d**, Forced expression of MITF or TFE3 in HPDE cells causes a decrease in p-ACC and p-AMPK levels. Error bars represent mean \pm s.d. Data are representative of at least 3 independent experiments. Significance was analysed using two-tailed Student's t -test.



Extended Data Figure 9 | Regulation of *in vitro* and *in vivo* growth by MiT/TFE proteins. **a**, A panel of human PDA cell lines were infected with shRNAs targeting MITF, TFE3 or TFE3. Growth relative to cells infected with shGFP control was assayed 8–10 days after infection. Note that sensitivity to individual knockdown correlates with the relative expression of each protein (see Extended Data Fig. 1f). Coloured bars indicate knockdown condition, which leads to the greatest growth impairment. **b**, Selective sensitivity of PDA cells compared with non-PDA cells (QGP1, HPNE and the non-small-cell lung cancer cell line H460) to treatment with 50 μ M chloroquine (CQ) for 4 days. Error bars indicate mean \pm s.e.m. $*P < 0.05$. **c**, Immunoblot showing robust detection of LC3-II across a panel of primary patient-derived cultures (PDAC 1–6) and PDA cell lines. **d**, Subcellular fractionation showing that PDA patient cultures have constitutively nuclear MITF and TFE3. **e**, qRT-PCR showing that MITF (left) or TFE3 (right) knockdown suppressed multiple autophagy-lysosome genes in patient-derived PDA cells. $*P < 0.01$. **f**, Immunofluorescence staining for LAMP2 showing that MITF ($N = 212$

siCTRL, $N = 220$ siMITF), TFE3 ($N = 228$ siCTRL, $N = 244$ siTFE3) and TFE3 ($N = 229$ siCTRL, $N = 271$ siTFE3) knockdown results in enlarged, dysmorphic lysosomes in patient-derived PDA cultures and quantified in the graph on the right. Scale bar, 7.5 μ m. $**P < 0.0001$. **g**, Knockdown of the indicated MiT/TFE factors inhibits colony formation in a series of primary PDA cultures. **h**, 8988T cells infected *in vitro* with shGFP or shTFE3 (left) and PL18 cells infected with shGFP or two hairpins targeting MITF (shMITF_1 and shMITF_2; right) were implanted subcutaneously on both flanks of SCID mice ($N = 4$ mice per group). Tumour xenograft growth was monitored over the course of 70 (8988T) and 50 (PL18) days. Error bars indicate mean \pm s.e.m. **i**, Forced expression of MITF in *Kras*^{G12D} mouse PanIN cells causes an increase in autophagy-lysosome gene expression relative to control cells *in vitro*. $*P < 0.005$. Data are representative of $N = 3$ independent experiments. Significance was analysed using two-tailed Student's *t*-test.

Energetic coupling between plastids and mitochondria drives CO₂ assimilation in diatoms

Benjamin Bailleul^{1,2,3,4}, Nicolas Berne¹, Omer Murik⁴, Dimitris Petroutsos⁵, Judit Prihoda⁴, Atsuko Tanaka⁴, Valeria Villanova⁶, Richard Bligny⁵, Serena Flori⁵, Denis Falconet⁵, Anja Krieger-Liszkay⁷, Stefano Santabarbara⁸, Fabrice Rappaport³, Pierre Joliot³, Leila Tirichine⁴, Paul G. Falkowski², Pierre Cardol¹, Chris Bowler⁴ & Giovanni Finazzi⁵

Diatoms are one of the most ecologically successful classes of photosynthetic marine eukaryotes in the contemporary oceans. Over the past 30 million years, they have helped to moderate Earth's climate by absorbing carbon dioxide from the atmosphere, sequestering it via the biological carbon pump and ultimately burying organic carbon in the lithosphere¹. The proportion of planetary primary production by diatoms in the modern oceans is roughly equivalent to that of terrestrial rainforests². In photosynthesis, the efficient conversion of carbon dioxide into organic matter requires a tight control of the ATP/NADPH ratio which, in other photosynthetic organisms, relies principally on a range of plastid-localized ATP generating processes^{3–6}. Here we show that diatoms regulate ATP/NADPH through extensive energetic exchanges between plastids and mitochondria. This interaction comprises the re-routing of reducing power generated in the plastid towards mitochondria and the import of mitochondrial ATP into the plastid, and is mandatory for optimized carbon fixation and growth. We propose that the process may have contributed to the ecological success of diatoms in the ocean.

In oxygenic photosynthesis, light drives a linear electron flow from water to NADPH by the two photosystems (PS I and PS II), and the generation of an electrochemical proton gradient (or proton motive force, PMF) across the thylakoid membranes, which fuels ATP synthesis by an ATP synthase. Although the ratio of ATP/NADPH generated by linear electron flow is not entirely resolved^{7,8}, it is considered to be insufficient to fuel CO₂ import into the plastid and assimilation by the Calvin cycle^{8,9}. Therefore, to make up the shortfall, additional ATP must be produced by alternative pathways that do not generate NADPH. In Viridiplantae (including green algae and higher plants) these alternative electron pathways have been found in the chloroplast and mostly comprise cyclic electron flow (CEF) around PS I (ref. 3) and/or the water-to-water cycles¹⁰; that is, flows of electrons resulting from the oxidation of water at PS II that are re-routed to an oxidase activity. This last group of oxidases include the Mehler reaction at the PS I acceptor side^{4,11,12}, the activity of the plastoquinone terminal oxidase downstream of PS II (ref. 5) and the oxygenase activity of ribulose 1,5-bisphosphate carboxylase–oxygenase (Rubisco) (photorespiration⁶). Although genes encoding most components for these processes appear to be present in diatoms^{13–15}, it is currently unknown what mechanisms are used to balance the ATP/NADPH ratio. We therefore investigated this question using the model species *Phaeodactylum tricornutum*.

The PMF generated across thylakoid membranes comprises an electric field ($\Delta\psi$) and a proton gradient (ΔpH). The $\Delta\psi$ can be probed

in vivo by measuring the electro-chromic shift (ECS), that is the Stark effect, a modification of the absorption spectrum of specific pigments caused by changes in the transmembrane electric field in the plastid¹⁶. An ECS signal is present in *P. tricornutum* (Fig. 1a), and an analysis of the ECS signal relaxation after light exposure (Extended Data Fig. 1) reveals that it comprises two components displaying different spectra (Fig. 1a). One follows a linear dependence on the amplitude of $\Delta\psi$ whereas the other follows a quadratic relationship (Fig. 1b). The existence of a 'quadratic ECS' is predicted by theory¹⁶ but has only been observed so far in mutants of green algae with altered pigment composition¹⁷. The peculiar existence of two different ECS probes in wild-type *P. tricornutum* cells allows an absolute quantification of the electric field, providing a valuable tool to analyse the PMF in a living cell (see Methods).

We plotted the amplitude of the quadratic versus linear ECS signals during the relaxation of a light-induced PMF and obtained a parabolic function (Fig. 1c, d and Extended Data Fig. 2). However, the ECS signals did not reach the minimum of the parabola in the dark, but rather remained positive. This indicates that a PMF is maintained across the thylakoid membrane of diatoms in the dark ($\Delta\psi_{\text{d}}$, Fig. 1c). The field $\Delta\psi_{\text{d}}$ can not only be dissipated by addition of an uncoupler (carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP)), but also by blocking mitochondrial electron transport by anaerobiosis or addition of antimycin A (AA) plus salicylhydroxamic acid (SHAM). The combination of these two inhibitors blocks both the cyanide-sensitive (complex III) and the cyanide-insensitive (alternative oxidase, AOX) respiratory pathways (Fig. 1d). These results suggest that the residual PMF in the dark is generated in plastids by the chloroplast ATPase by hydrolysis of ATP derived from mitochondria (Fig. 1e)¹⁸. Furthermore, the extent of $\Delta\psi_{\text{d}}$ observed in *P. tricornutum* is larger than that previously reported in green algae¹⁹, suggesting that the ATP exchange could be more efficient in diatoms.

To evaluate what mechanism regulates ATP/NADPH in the light in *P. tricornutum*, we first used the linear ECS to probe the CEF capacity (see Methods). CEF turned out to represent only a very low fraction of the maximum electron flow capacity (Fig. 2a and Extended Data Fig. 3a, c) and was insensitive to changes in the photosynthetic flux (Fig. 2a). Thus, it appears very unlikely that CEF could regulate ATP/NADPH fluxes. Next we explored the water-to-water cycle using membrane-inlet mass spectrometry on cells incubated with ¹⁸O₂. O₂ consumption increased with light, being ~2.5-fold higher at saturating light intensities than in the dark (Extended Data Fig. 3b, d). We also found that the light-stimulated O₂ consumption was blocked by 3-(3,4-dichlorophenyl)-1,1-dimethyl-urea (DCMU), which inhibits

¹Génétique et Physiologie des Microalgues, Département des Sciences de la vie et PhytoSYSTEMS, Université de Liège, B-4000 Liège, Belgium. ²Environmental Biophysics and Molecular Ecology Program, Departments of Marine and Coastal Sciences and of Earth and Planetary Sciences, Rutgers University, New Brunswick, New Jersey 08901, USA. ³Institut de Biologie Physico-Chimique (IBPC), UMR 7141, Centre National de la Recherche Scientifique (CNRS), Université Pierre et Marie Curie, 13 Rue Pierre et Marie Curie, F-75005 Paris, France. ⁴Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France. ⁵Laboratoire de Physiologie Cellulaire et Végétale, UMR 5168, Centre National de la Recherche Scientifique (CNRS), Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Université Grenoble Alpes, Institut National Recherche Agronomique (INRA), Institut de Recherche en Sciences et Technologies pour le Vivant (IRTSV), CEA Grenoble, F-38054 Grenoble cedex 9, France. ⁶Fermentalg SA, F-33500 Libourne, France. ⁷Institute for Integrative Biology of the Cell (I2BC), Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Centre National de la Recherche Scientifique (CNRS), Université Paris-Sud, Institut de Biologie et de Technologie de Saclay, F-91191 Gif-sur-Yvette cedex, France. ⁸Istituto di Biofisica, Consiglio Nazionale delle Ricerche, Via Celoria 26, I-20133 Milan, Italy.

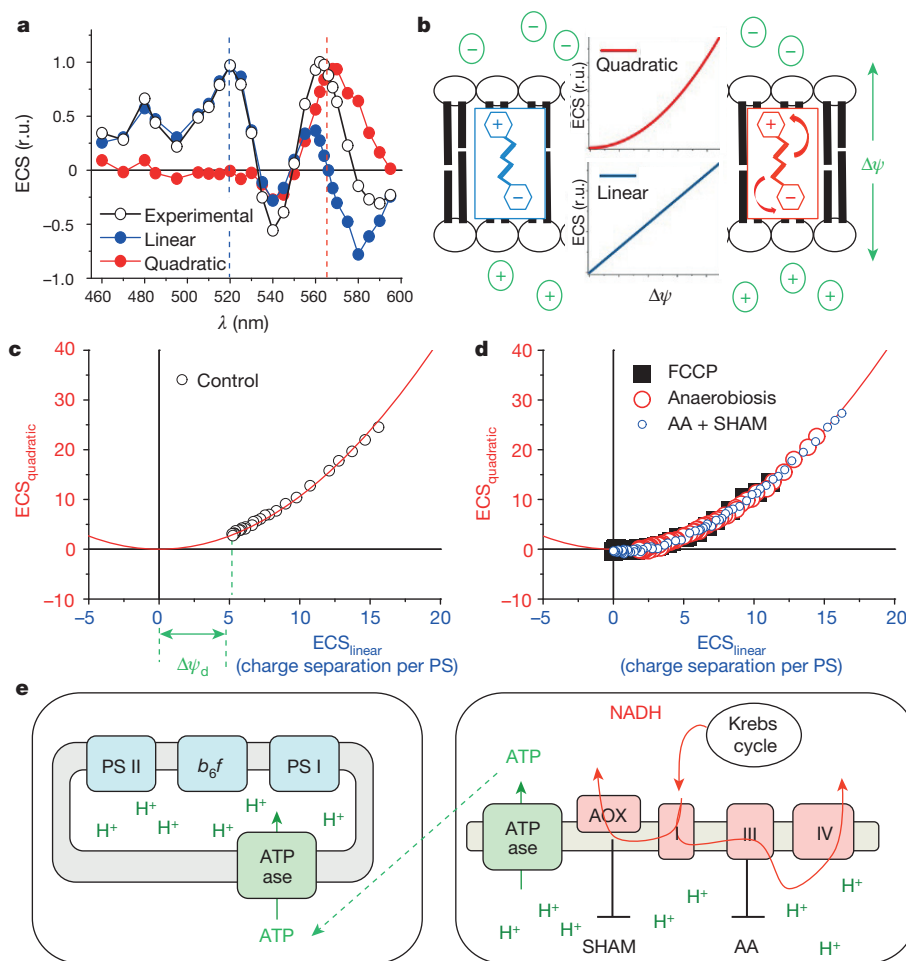


Figure 1 | ECS measures the PMF in *P. tricornutum*. **a**, Deconvolution of the experimental ECS spectrum (black) into linear (blue) and quadratic (red) spectral components (see Methods); r.u., relative units. **b**, Schematic representation of polar (blue) and polarizable (red) pigments, and their associated linear (blue) and quadratic (red) ECS responses to the electric field. Black: thylakoid lipid bilayer. Green '+' and '-' symbols: $\Delta\psi$. Red arrows: pigment polarization induced by $\Delta\psi$. **c**, **d**, Relationship between quadratic and linear ECS in control (**c**) and in cells treated with uncoupler (8 μ M FCCP, black squares), anaerobiosis (red circles) and respiratory inhibitors (AA, 5 μ M, and SHAM, 1 mM; blue circles) (**d**). Green arrow: extent of the dark electric field ($\Delta\psi_d$). Representative of five (**c**) and three (**d**) independent biological samples. **e**, Schematic representation of the energetic interactions between plastid (left) and mitochondria (right) in the dark. Red arrows: respiratory electron flows. Green dashed line: putative ATP/ADP exchange pathway between the organelles. ATPase, ATPase/synthase; b_6f , cytochrome b_6f ; I/III/IV, respiratory complexes I, III and IV.

O_2 production by PS II (Extended Data Fig. 3b, d), indicating that this process is fed by electrons generated by PS II. O_2 consumption increased linearly with O_2 production, in agreement with earlier findings in another diatom species²⁰, indicating that a constant proportion (~10%) of the electron flow from photosynthesis is re-routed to an O_2 -consuming pathway, regardless of light intensity (Fig. 2b).

To test whether the O_2 -consuming pathway occurs in the plastid or relies on mitochondrial activity, we used increasing concentrations of inhibitors to titrate respiration and tested possible consequences on photosynthesis. We reasoned that if respiration consumes reducing equivalents generated in the plastid to generate additional ATP, any mitochondrial dysfunction should negatively impact photosynthesis.

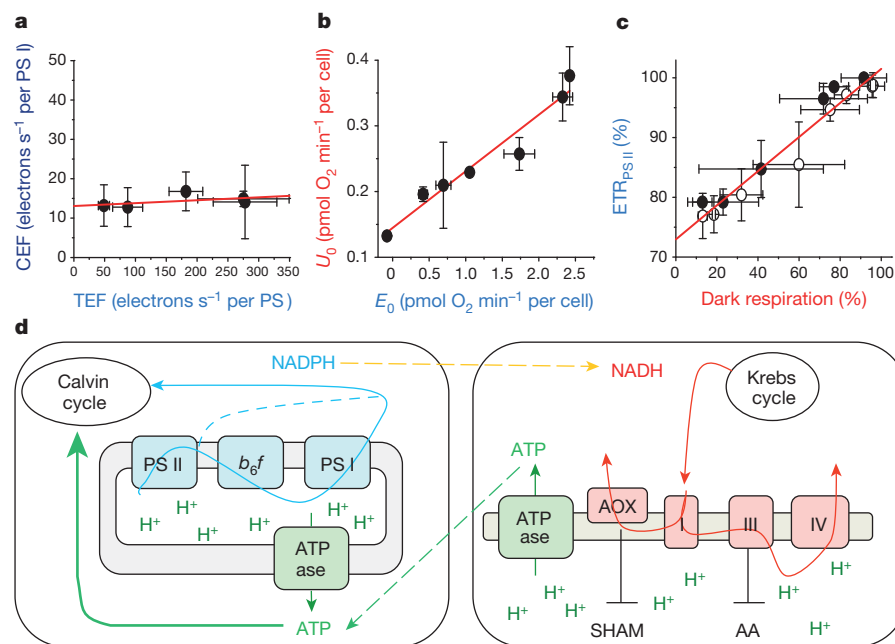


Figure 2 | Energetic interactions between mitochondria and plastid in *P. tricornutum*. **a**, Relationship between CEF capacity and total electron flow (TEF, mean \pm s.d. from data in Extended Data Fig. 3a, c). **b**, Relationship between oxygen uptake (U_0) and gross photosynthesis (E_0) as measured by membrane-inlet mass spectrometry (mean \pm s.e.m. from data in Extended Data Fig. 3b, d). **c**, Dependency of photosynthetic activity ($ETR_{ps II}$) on respiration rates (as the percentage of control, mean \pm s.d. from data in Extended Data Fig. 4). Closed circles: SHAM + AA; open circles: SHAM + myxothiazol treatments (see Methods). **d**, Schematic representation of possible plastid-mitochondria metabolic interactions in the light. Continuous and dashed blue arrows: photosynthetic linear and cyclic flows, respectively. Yellow arrow: exchange of reducing equivalents. For other symbols, see text and Fig. 1e.

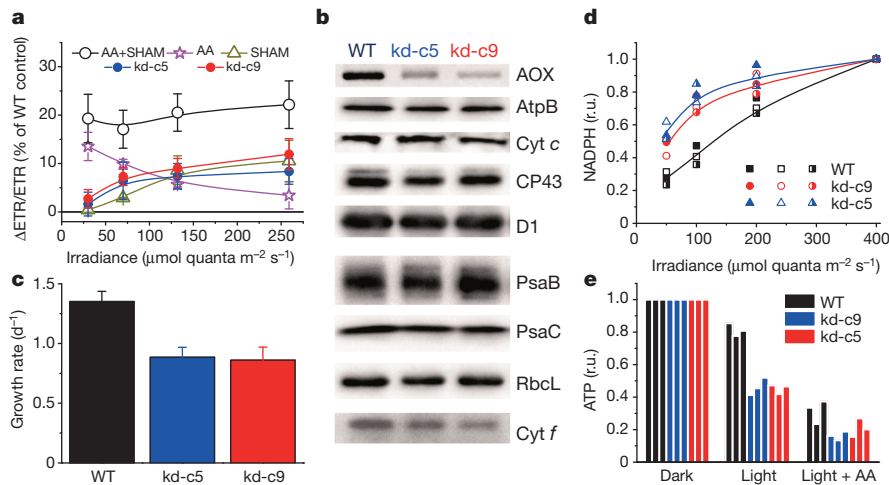


Figure 3 | Phenotypic traits of AOX knockdown lines of *P. tricornutum*. **a**, Relative sensitivity of photosynthesis ($ETR_{PS II}$) to addition of respiratory inhibitors: AA (magenta), SHAM (dark yellow) and AA + SHAM (black) ($n = 3 \pm S.D.$), or to knockdown of AOX ($n = 5 \pm S.D.$). Blue and red: kd-c5 and kd-c9, respectively. **b**, Western blot analysis of photosynthetic and respiratory complexes. **c**, Growth rates of the wild-type and AOX lines ($n = 7 \pm S.D.$). **d**, *In vivo* light dependency of NADPH redox state in wild-type and AOX lines. Data are normalized to the maximum value in the light. **e**, *In vivo* ^{31}P -NMR evaluation of the NTP content in wild-type and AOX knockdown lines, in the dark or in low light, with or without AA (data normalized to the dark values). **d**, **e**, Three independent biological samples.

We found that this was indeed the case, as photosynthetic electron transfer rate ($ETR_{PS II}$) linearly followed changes in respiration (Fig. 2c and Extended Data Fig. 4). We conclude that a partial re-routing of the photosynthetic flow towards mitochondrial respiration rather than CEF optimizes photosynthesis in diatoms, providing commensurate ATP per NADPH at all irradiances (Fig. 2d).

Photosynthetic activity displayed increasing sensitivity to AOX inhibition with light, suggesting that cyanide-insensitive respiration becomes prominent in high light (Fig. 3a). This prompted us to generate AOX knockdown cell lines of *P. tricornutum*. Two independent clones were selected on the basis of reduced AOX protein accumulation (Fig. 3b) and decreased activity (measured as the SHAM-sensitive, AA-insensitive component of respiration; Extended Data Fig. 5a). The AOX contribution, representing $\sim 50\%$ of dark respiration in wild-type cells (Extended Data Fig. 4e), was decreased two-fold in the two knockdown lines (Extended Data Fig. 5a). Confocal microscopy confirmed the mitochondrial localization of the targeted gene product (Extended Data Fig. 6a). The reduced AOX activity in the knockdown lines paralleled a diminished PMF in the dark ($\Delta\Psi_d$, Extended Data

Fig. 5b), despite the fact that overall dark respiration was slightly higher (Extended Data Fig. 5a). This effect was strongly enhanced by addition of AA (Extended Data Fig. 5b). The decreased AOX activity also correlated with a decreased photosynthetic capacity, especially under high light intensities (similar to SHAM-treated wild-type cells; Fig. 3a), and a diminished growth rate (Fig. 3c), which was exacerbated further by inhibiting complex III using AA (Extended Data Fig. 5c). The growth and photosynthetic phenotypes were not due to changes in the accumulation of the photosynthetic complexes, for which we detected comparable levels of representative proteins in all cell lines (Fig. 3b). The only exception was a small decrease in the cytochrome b_6f content in the knockdown cell lines, which nonetheless did not decrease its overall catalytic turnover (Extended Data Fig. 7).

Our working model presented in Fig. 2d predicts that disruption of the plastid-mitochondria interaction in the knockdown cell lines should lead to the accumulation of NADPH and a decreased cellular content of ATP in the light. Indeed, *in vivo* assessments of the pools of NADPH and ATP in wild-type and knockdown cell lines confirmed an increase in the NADPH/NADP $^{+}$ ratio with light intensity

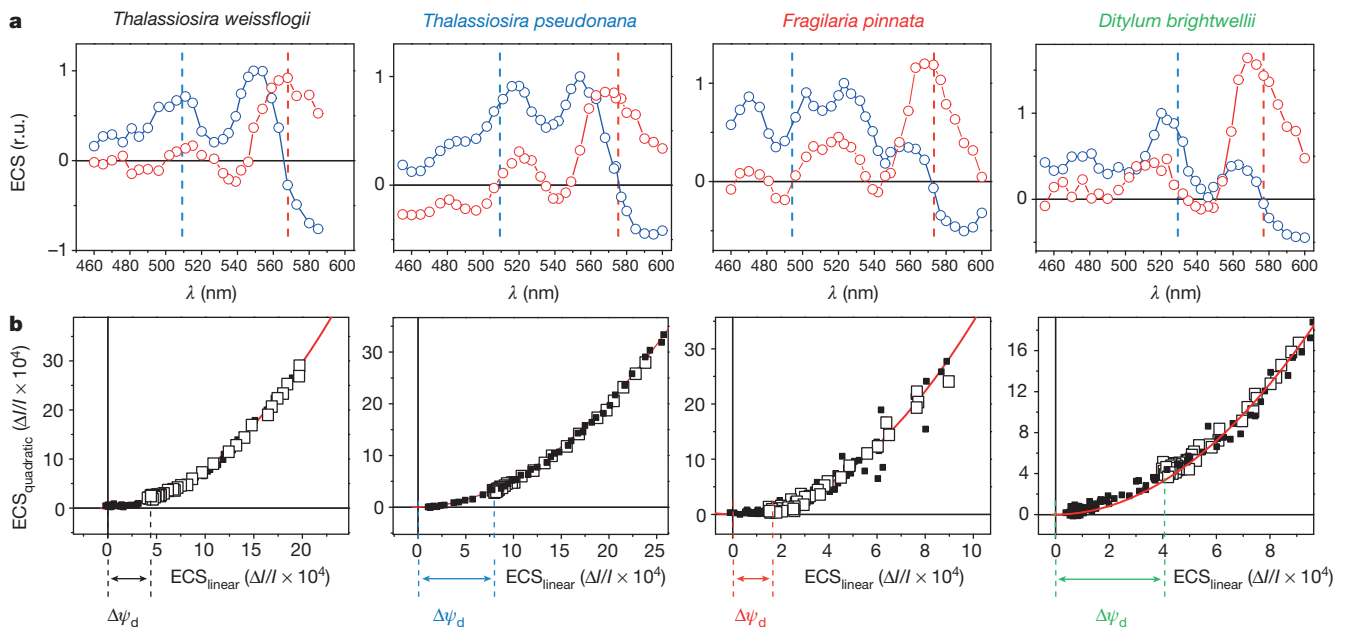


Figure 4 | ATP transfer from mitochondria to plastid in representative diatoms. **a**, Spectra of the linear (blue) and quadratic (red) ECS probes in four diatoms. Blue and red vertical dashed lines represent the wavelengths used for linear and quadratic ECS, respectively. Spectra are normalized to 1 at the

maximum value of the linear ECS. **b**, Relationship between the quadratic and the linear ECS in control conditions (open symbols) and in AA + SHAM conditions (filled symbols). $\Delta\Psi_d$ is represented as a horizontal arrow. Data are representative of three independent biological samples.

(Fig. 3d and Extended Data Fig. 8a), accompanied by a net decrease of cellular ATP levels (Fig. 3e and Extended Data Fig. 8b), both of which were more drastic in AOX knockdown cells than wild-type cells. These observations confirm that mitochondrial respiration is directly involved in the adjustment of the ATP to NADPH ratio in the plastid.

We then examined the generality of our findings in other diatom species. The similar ECS features (linear and quadratic components) in *Thalassiosira pseudonana*, *Thalassiosira weissflogii*, *Fragilaria pinnata* and *Ditylum brightwellii* (Fig. 4a) were used to confirm the presence of a PMF in the plastids in the dark at the expense of hydrolysis of ATP supplied by the mitochondria in all cases (Fig. 4b). Moreover, a negligible contribution of CEF (Extended Data Fig. 9) and a significant involvement of mitochondrial respiration to photosynthesis (Extended Data Fig. 10) were found in all these species. The involvement of mitochondrial respiration in the optimization of photosynthesis therefore appears to be a general and conserved feature in diatoms.

We conclude that ATP generation sets the rate of photosynthetic carbon assimilation in diatoms, as suggested in other photosynthetic organisms⁸, but that in contrast with the Viridiplantae, optimization of diatom photosynthesis does not rely on plastid-localized processes. Instead, constitutive energetic interactions between diatom mitochondria and plastids ensure the sharing of reducing equivalents and ATP to fuel CO₂ assimilation in the light (Fig. 2d). While the process we have uncovered has some similarities to the export of reducing equivalents from the plastids towards mitochondria in plants and green algae, the fundamental difference is that in plants and green algae the process serves as a valve to dissipate excess of reducing power²¹, and can only participate in the regulation of the ATP/NADPH ratio when the chloroplast capacity to make extra ATP is genetically disrupted^{22–24}. We propose that the presence of transporters such as the malate shuttle²¹, which are encoded in diatom genomes¹³, as well as the very tight physical interactions observed in diatoms between plastids and mitochondria (Extended Data Fig. 6b), may make these energetic interactions possible between the two organelles. Because diatom plastids are surrounded by four membranes rather than two as in Viridiplantae¹³, it will be of interest to elucidate the configuration of such transporters, as well as the rationale for the diminished CEF efficiency in diatoms. More generally, the coupling of respiratory and photosynthetic activities in diatoms should be explored in the context of resource utilization in the ocean and as a means to boost the production of useful metabolites for biotechnology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 January; accepted 28 May 2015.

Published online 13 July; corrected online 19 August 2015 (see full-text HTML version for details).

1. Falkowski, P. G. The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. G. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
3. Shikanai, T. Cyclic electron transport around photosystem I: genetic approaches. *Annu. Rev. Plant Biol.* **58**, 199–217 (2007).
4. Asada, K. The water–water cycle as alternative photon and electron sinks. *Phil. Trans. R. Soc. Lond. B* **355**, 1419–1431 (2000).
5. Cardol, P. *et al.* An original adaptation of photosynthesis in the marine green alga *Ostreococcus*. *Proc. Natl Acad. Sci. USA* **105**, 7881–7886 (2008).

6. Ort, D. R. & Baker, N. R. A photoprotective role of O₂ as an alternative electron sink in photosynthesis? *Curr. Opin. Plant Biol.* **5**, 193–198 (2002).
7. Petersen, J., Förster, K., Turina, P. & Gräber, P. Comparison of the H⁺/ATP ratios of the H⁺-ATP synthases from yeast and from chloroplast. *Proc. Natl Acad. Sci. USA* **109**, 11150–11155 (2012).
8. Allen, J. F. Photosynthesis of ATP-electrons, proton pumps, rotors, and poise. *Cell* **110**, 273–276 (2002).
9. Lucker, B. & Kramer, D. M. Regulation of cyclic electron flow in *Chlamydomonas reinhardtii* under fluctuating carbon availability. *Photosynth. Res.* **117**, 449–459 (2013).
10. Allen, J. F. Oxygen reduction and optimum production of ATP in photosynthesis. *Nature* **256**, 599–600 (1975).
11. Radmer, R. J. & Kok, B. Photoreduction of O₂ primes and replaces CO₂ assimilation. *Plant Physiol.* **58**, 336–340 (1976).
12. Badger, M. R. Photosynthetic oxygen exchange. *Annu. Rev. Plant Physiol.* **36**, 27–53 (1985).
13. Prihoda, J. *et al.* Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* **63**, 1543–1557 (2012).
14. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).
15. Grouneva, I., Rokka, A. & Aro, E.-M. The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosynthetic machinery. *J. Proteome Res.* **10**, 5338–5353 (2011).
16. Witt, H. T. Energy conversion in the functional membrane of photosynthesis. Analysis by light pulse and electric pulse methods. The central role of the electric field. *Biochim. Biophys. Acta* **505**, 355–427 (1979).
17. Joliot, P. & Joliot, A. Characterization of linear and quadratic electrochromic probes in *Chlorella sorokiniana* and *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta* **975**, 355–360 (1989).
18. Diner, B. & Joliot, P. Effect of the transmembrane electric field on the photochemical and quenching properties of photosystem II *in vivo*. *Biochim. Biophys. Acta* **423**, 479–498 (1976).
19. Finazzi, G. & Rappaport, F. *In vivo* characterization of the electrochemical proton gradient generated in darkness in green algae and its kinetics effects on cytochrome b₆f turnover. *Biochemistry* **37**, 9999–10005 (1998).
20. Waring, J., Klenell, M., Bechtold, U., Underwood, G. J. C. & Baker, N. R. Light-induced responses of oxygen photo-reduction, reactive oxygen species production and scavenging in two diatom species. *J. Phycol.* **46**, 1206–1217 (2010).
21. Kinoshita, H. *et al.* The chloroplastic 2-oxoglutarate/malate transporter has dual function as the malate valve and in carbon/nitrogen metabolism. *Plant J.* **65**, 15–26 (2011).
22. Lemaire, C., Wollman, F. A. & Bennoun, P. Restoration of phototrophic growth in a mutant of *Chlamydomonas reinhardtii* in which the chloroplast atpB gene of the ATP synthase has a deletion: an example of mitochondria-dependent photosynthesis. *Proc. Natl Acad. Sci. USA* **85**, 1344–1348 (1988).
23. Cardol, P. *et al.* Impaired respiration discloses the physiological significance of state transitions in *Chlamydomonas*. *Proc. Natl Acad. Sci. USA* **106**, 15979–15984 (2009).
24. Dang, K. V. *et al.* Combined increases in mitochondrial cooperation and oxygen photoreduction compensate for deficiency in cyclic electron flow in *Chlamydomonas reinhardtii*. *Plant Cell* **26**, 3036–3050 (2014).

Acknowledgements This work was supported by grants from Agence Nationale de la Recherche (ANR-12-BIME-0005, DiaDomOil to C.B., D.P. and G.F.; ANR-8NT09567009, Phytadapt to B.B., G.F. and C.B.; ANR-11-LABX-0011-01, Dynamo to F.R. and P.J.; ANR-11-IDEX-0001-02, PSL Research University and ANR-10-LABX-54, MEMOLIFE to C.B.), the Région Rhône-Alpes (Cible project) to G.F., the Marie Curie Initial Training Network Accliphot (FP7-PEOPLE-2012-ITN; 316427) to G.F., D.P., S.F. and V.V., an ERC Advanced Award (Diatomite) and the EU MicroB3 project to C.B., the CNRS Défi (ENRS 2013) to G.F. and L.T., and the CEA Bioénergies program to G.F. and D.P. P.C., N.B. and B.B. acknowledge financial support from the Belgian Fonds de la Recherche Scientifique F.R.S.-F.N.R.S. (F.R.F.C. 2.4597.11, CDR J.0032.15 and Incentive Grant for Scientific Research F.4520). B.B. also acknowledges a post-doctoral fellowship from Rutgers University and J.P. was funded from the COSI ITN project to C.B. Thanks are due to J.-L. Putaux and C. Lancelon-Pin for help with electron microscopy, to L. Moyet for technical support for the *in vivo* NMR analysis, to A. E. Allen for the AOX antibody, and to A. Falcitatore and F. Barneche for critical reading the manuscript.

Author Contributions B.B., L.T., C.B. and G.F. designed the study. B.B., N.B., O.M., D.P., J.P., A.T., V.V., R.B., S.F., D.F., A.K.-L., F.R., P.J., L.T., P.C. and G.F. performed experiments. B.B., N.B., O.M., D.P., R.B., A.K.-L., S.S., F.R., P.J., L.T., P.F., P.C., C.B. and G.F. analysed the data. B.B., C.B. and G.F. wrote the manuscript, and all authors revised and approved it.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.B. (bailleul@ibpc.fr), C.B. (cbowler@biologie.ens.fr) or G.F. (giovanni.finazzi@cea.fr).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Growth conditions. Wild-type and AOX transformant lines of *P. tricornutum* Pt1 8.6 (CCMP 2561) were grown in artificial sea water (ASW²⁵). *T. pseudonana* (CCMP 1335), *T. weissflogii* (CCMP 1336), *F. pinnata* (CCAP 1029/2) and *D. brightwellii* (CCMP 359) were grown in F/2 medium, supplemented with silicate²⁶. All strains were grown at $19 \pm 1^\circ\text{C}$ in semi-continuous batch culture (with moderate shaking for *P. tricornutum*, *T. pseudonana* and *T. weissflogii*). The photoperiod was 12 h light/12 h dark, and light irradiance was $70 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$. Cell concentration was determined daily with a Z2 Coulter Counter analyser (Beckman Coulter) to ensure all the experiments were performed with cells in exponential phase. For biophysical measurements, cells were concentrated by centrifugation and resuspended in their growth medium (supplemented with 10% w/v Ficoll to prevent cell sedimentation) and kept in the dark at least 30 min before measurements.

Inhibitors. DCMU, 2,5-dibromo-3-methyl-6-isopropyl-p-benzoquinone (DBMIB), FCCP, AA, myxothiazol and SHAM (Sigma-Aldrich) were dissolved in ethanol, whereas hydroxylamine, glucose, glucose oxidase and catalase (Sigma-Aldrich) were dissolved in deionized water. FCCP was used at a very low concentration ($8 \mu\text{M}$, Fig. 1d) to allow the disruption of the dark PMF without preventing the light-induced generation of PMF needed to quantify $\Delta\Psi_d$. AA and myxothiazol were used at $5 \mu\text{M}$, unless otherwise stated. DCMU was used at a concentration of $15 \mu\text{M}$. In every measurement involving hydroxylamine or SHAM, the lowest inhibitor concentration to induce a full inhibition of PS II activity or maximum inhibition of respiration, respectively, was used. The range of concentrations used was $30\text{--}100 \mu\text{M}$ and $0.5\text{--}1 \text{ mM}$ for hydroxylamine and SHAM, respectively. Anaerobic conditions were obtained through incubation with catalase ($1,000 \text{ U ml}^{-1}$), glucose (10 mM) and glucose oxidase ($20,000 \text{ U ml}^{-1}$). AA has been previously described as an inhibitor of cyclic electron flow, affecting the NDH-independent pathway²⁷. This potential effect was ruled out in diatoms because no change in cyclic electron flow was noticed upon addition of AA. Consistent with the fact that genes encoding some members of the NDH complex are absent in diatom genomes, this indicates that AA does not affect the NDH-independent CEF pathway in diatoms. AA and myxothiazol were preferred to potassium cyanide to block the cyanide sensitive pathway of respiration because potassium cyanide also affects Rubisco activity²⁸, ascorbate peroxidase²⁹ and Cu/Zn superoxide dismutase³⁰.

Deconvolution of linear and quadratic ECS components. To deconvolute the linear and quadratic contributions to the ECS signals, cells were left for an hour in the cuvette to reach complete anaerobiosis. In these conditions, the ATP synthase activity is slowed down³¹, and long-living ECS signals are no longer contaminated with other light-induced absorption changes (principally associated with *c*-type cytochromes). Light stimulation of cells was achieved with a series of six laser single-turnover (duration $\sim 7 \text{ ns}$) saturating flashes, provided by a laser dye (LDS 698) pumped by a frequency doubled Nd-YAG laser (Quintel). We considered that the relaxation of the electric field generated by the light stimulus is described by the exponential function $\Delta\Psi = \Delta\Psi_0 \exp(-t/\tau)$, where t is time, $\Delta\Psi_0$ is the initial electric field generated by the light, and τ is the electric-field decay lifetime. The linear and quadratic components of the ECS are theoretically proportional to $\Delta\Psi$ and $\Delta\Psi^2$, respectively. Therefore, the $\Delta I/I$ spectro-temporal matrices (from 460 to 600 nm) can be described by a sum of two exponentials: $y(\lambda, t) = A(\lambda) \exp(-t/\tau) + B(\lambda) \exp(-(2t/\tau)) + C(\lambda)$. The kinetics of ECS relaxation were fitted by a global routine, which considers the lifetime τ as a global (wavelength-independent) variable, and the amplitudes of linear and quadratic components (A and B , respectively) as local (wavelength-dependent) variables. A non-decaying component (C) was also included in the fit to account for a small fraction of residual signal at long delay times. The plot of the A and B amplitudes as a function of the wavelength provides the decay associated spectra of the linear and quadratic contributions to the ECS signal, respectively, which are shown in Fig. 1a. The fit was performed with homemade software, which used the MINUIT package, developed and distributed by CERN (Geneva, Switzerland) and implemented in FORTRAN77. It minimizes the reduced sum of squared residues between the model function and the experimental data, employing a two-step protocol involving an initial search that utilizes the Simplex method (Nelder–Mead algorithm) and a refined search using the Levenberg–Marquardt algorithm as described in ref. 32. The quality of the fit description was judged on the basis of reduced sum of squared residues statistics, visual inspection of the fit residuals, residuals autocorrelation and stability of the solutions upon random perturbation of the best-fit. The deconvolution was performed on two independent biological samples, giving similar results.

Measurements of *c*-type cytochromes and linear and quadratic ECS.

Absorption difference signals were measured at different wavelengths with a Joliot-type spectrophotometer (JTS-10, Biologic), equipped with a white probing LED and the appropriate interference filters (3–8 nm bandwidth). For $\Delta\Psi_d$ measurements, the PMF was increased using an $\sim 10 \text{ ms}$ pulse of saturating ($4,500 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$) red light (see Extended Data Fig. 3 for representative ECS kinetics). For *P. tricornutum*, ECS signals were evaluated using three wavelengths, to eliminate contribution from *c*-type cytochromes (see Extended Data Fig. 2). The latter was calculated as $\text{cyt } c = [554] - 0.4[520] - 0.4[566]$, where $[554]$, $[520]$ and $[566]$ are the absorption difference signals at 554 nm, 520 nm and 566 nm, respectively. The very similar relaxation of *c*-type cytochromes in aerobic and anaerobic conditions, despite very different ECS relaxations (Extended Data Fig. 2b), demonstrates the validity of the $\text{cyt } c$ deconvolution procedure. Then, ecs_{lin} and ecs_{quad} (ECS signals before correction for $\Delta\Psi_d$) were estimated from the following relationships: $\text{ecs}_{\text{lin}} = [520] - 0.25\text{cyt } c$ and $\text{ecs}_{\text{quad}} = [566] + 0.15\text{cyt } c$. For the other diatoms, appropriate wavelengths were chosen for calculating ecs_{lin} and ecs_{quad} (red and blue lines in Fig. 4) to minimize the cytochrome *c* contributions. The relationships between ecs_{quad} and ecs_{lin} were fitted with the parabolic equation $\text{ecs}_{\text{quad}} + a\Delta\Psi_d^2 = a(\text{ecs}_{\text{lin}} + \Delta\Psi_d)^2$, where $\Delta\Psi_d$ is the electrical component of the PMF in the dark and a is constant for all the conditions in a diatom species (see Extended Data Fig. 2c). The ecs_{lin} and ecs_{quad} values represent ECS changes relative to dark values. Therefore, we corrected them for the dark electric field. This leads to $\text{ECS}_{\text{lin}} = \text{ecs}_{\text{lin}} + \Delta\Psi_d$ and $\text{ECS}_{\text{quad}} = \text{ecs}_{\text{quad}} + a\Delta\Psi_d^2$, namely absolute values of the ECS signals. This simply corresponds to a shift of the x - and y -axes to allow the minimum of the parabola to coincide with the origin of the axes (see Extended Data Fig. 2d), and gives $\text{ECS}_{\text{quad}} = a\text{ECS}_{\text{lin}}^2$. This leads to the evaluation of $\Delta\Psi_d$ as the minimal ECS_{lin} value of the experimental data. In Fig. 1c, d, and Extended Data Fig. 2c, d, ECS data were normalized to the ecs_{lin} increase upon a saturating laser flash (that is, one charge separation per photosystem, see ref. 33). This allows the estimation of $\Delta\Psi_d$ in *P. tricornutum* as approximately five charge separations by PS, namely $\sim 100 \text{ mV}$ (ref. 12) (Fig. 1c). The presence of linear and quadratic ECS components allows measurement of the absolute value of $\Delta\Psi$ in the dark ($\Delta\Psi_d$). Indeed the amplitude of the linear ECS response (ecs_{lin}) observed upon a light stimulus increasing $\Delta\Psi$ is constant, namely independent of the value of the field pre-existing the illumination ($\Delta\Psi_d$). Conversely, the amplitude of the quadratic ECS response (ecs_{quad}) is a function of the value of $\Delta\Psi_d$. Therefore, plotting the amplitude of the ecs_{quad} versus ecs_{lin} allows quantification of the absolute value of the electric field in the dark ($\Delta\Psi_d$).

Cyt b_6f turnover was measured through the slow phase (phase b^{31}) of the linear ECS, which reflects b_6f -catalysed charge transfer across the membranes, and through the reduction rate of the *c*-type cytochromes (c_6f), using the three-wavelengths deconvolution procedure described above. Measurements were performed after a saturating laser flash.

Measurements of photosynthetic flows. For calculation of the TEF (the sum of linear and cyclic electron flows) and CEF capacities, we measured the photochemical rates in the absence and presence, respectively, of DCMU. In brief, under steady-state illumination conditions, the ECS signal results from concomitant transmembrane potential generation by PS II, the cytochrome b_6f complex and PS I, and from transmembrane potential dissipation by the plastid ATP synthase. When light is switched off, PS activities stop immediately, while ATP synthase and cytochrome b_6f complex activities remain (transiently) unchanged. Therefore, the difference between the slopes of the linear ECS signal (ECS_{lin}) measured in the light and after the light is switched off ($S_D - S_L$) is proportional to the rate of PS I and PS II photochemistry (that is, to the rate of 'total' electron flow, Extended Data Fig. 3a). Because the linear ECS has been normalized to the amplitude of the linear ECS signal induced by a saturating laser flash³⁴ (see above), the difference of slopes evaluates the number of charge separations per photosystem and per second. The rate of CEF can be evaluated using the same approach under conditions where PS II activity is inhibited by DCMU, and dividing this slope by the linear ECS signal induced by a saturating laser flash in the presence of PS II inhibitors (one charge separation per PS I (ref. 34)). This was done using saturating concentrations of DCMU, which block PS II oxidation, and of hydroxylamine, to avoid charge recombination within PS II.

Fluorescence-based measurements. Fluorescence-based photosynthetic parameters were measured with a fluorescence imaging setup described in ref. 35. Photosynthetic electron transfer rate $\text{ETR}_{\text{PS II}}$ and NPQ were calculated, respectively, as $I(F_m' - F)/F_m'$ and $(F_m - F_m')/F_m'$, where F and F_m' are the steady-state and maximum fluorescence intensities in light-acclimated cells (respectively), F_m is the maximal fluorescence intensity in dark-adapted cells, and I is the light irradiance in $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ (refs 36, 37). In Fig. 2c and Extended Data Fig. 4a, d, the light irradiance is $30 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$. The light saturation curves of $\text{ETR}_{\text{PS II}}$ were fitted with the exponential rise function

$P = P_{\max}(1 - \exp(-E/E_k))$, where P_{\max} is the maximal photosynthetic electron transport rate and E_k is the optimal light. $\Delta ETR/ETR$ (Fig. 3a) was calculated as $(ETR_{\text{ref}} - ETR) \times 100/ETR_{\text{ref}}$, the reference being the value measured in wild-type cells in untreated conditions.

Membrane-inlet mass spectrometry measurements. Samples were introduced in a 3 ml thermostated cuvette, which was connected to a Quadrupole Mass Spectrometer (QMS 200, Pfeiffer Vacuum Prisma) by a stainless steel vacuum tube (0.125 inches) passing through a water trap filled with ethanol and dry ice. The sample was separated from the tube via a gas-permeable inlet system (polytetrafluoroethylene (PTFE) membrane). $^{18}\text{O}_2$ was added as a bubble to the algal suspension, and the bubble was removed before the experiment. The measurements of the partial pressures of $^{16}\text{O}_2$ ($p^{16}\text{O}_2$, $m/z = 32$), $^{18}\text{O}_2$ ($p^{18}\text{O}_2$, $m/z = 36$) and argon ($m/z = 40$) were performed after the cuvette was sealed. A blue light-emitting diode (LED) source was connected to the cuvette, and the light irradiance was manually adjustable in the 0 to $\sim 800 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ range. The temperature was kept at $19 \pm 1^\circ\text{C}$ in the cuvette during the experiment.

To calculate gross O_2 production (E_0) and uptake (U_0), respectively production and consumption by the cells, we adapted the equations from ref. 38:

$$U_0 = (\Delta[^{18}\text{O}]/\Delta t + k[^{18}\text{O}]) / (([^{18}\text{O}] + [^{16}\text{O}]) / [^{18}\text{O}])$$

$$E_0 = (\Delta[^{16}\text{O}]/\Delta t + k[^{16}\text{O}]) + U_0(([^{18}\text{O}] + [^{16}\text{O}]) / [^{16}\text{O}])$$

where k is the rate constant of O_2 decrease measured in the absence of algae. We normalized O_2 to Argon (a biologically inert gas with very similar solubility properties), which decreases the sensitivity of O_2 measurements to fluctuations by $\sim 80\%$ (ref. 39). The gas concentrations were calibrated by measuring the air-equilibrated O_2 concentration (stirring deionized water in the open cuvette for at least 5 h) and background O_2 (bubbling with N_2).

Respiration rates. Respiration rates were measured as O_2 exchange rates using a Clark-type oxygen electrode at 19°C (Hansatech Instruments). AOX capacity was measured as SHAM-sensitive respiration in conditions where the cyanide-sensitive pathway was inhibited beforehand (AA, 5 μM).

ATP/NADPH *in vivo* measurements. $\text{NADP}^+/\text{NADPH}$ redox changes were followed in living cells using a Dual-PAM (Walz). NADPH fluorescence was measured at 460 nm, upon excitation in the near ultraviolet. Chlorophyll *a* concentration was $\sim 5 \mu\text{g ml}^{-1}$. ATP content was measured using an *in vivo* ^{31}P -AMX 400 NMR spectrometer equipped with a 25-mm multinuclear probe tuned at 161.9 MHz, and a homemade lighting system, as described in ref. 40. The relative ATP content was estimated *in vivo* from the surface of α -, β - and γ -phosphorus resonance peaks corresponding to the three phosphates of NTPs, which dominate the NMR spectra with inorganic phosphate and polyphosphates⁴¹.

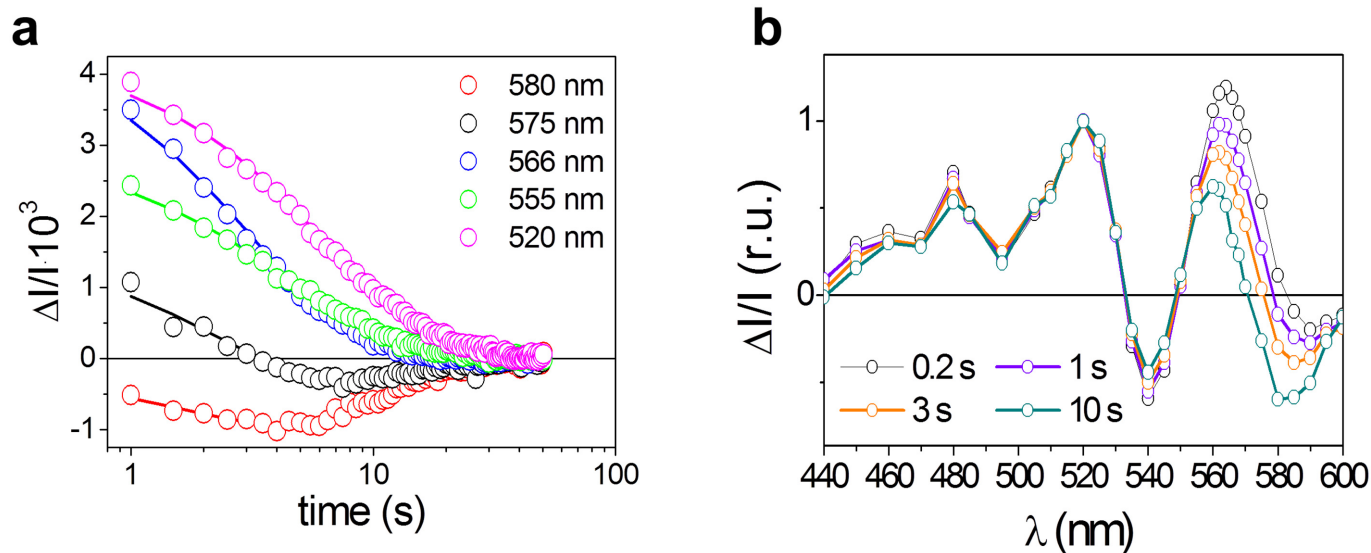
Western blots and immunolocalization. Protein samples (5–10 μg) were loaded on 13% SDS-polyacrylamide gel electrophoresis (PAGE) gels and blotted to nitrocellulose. Primary AOX antibody was custom designed (Sdix, 1:4,000 dilution). All other antisera used were obtained from Agrisera (<http://www.agrisera.com/en/info/catalog.html>). The blots were developed with ECL detection reagent and images of the blots were obtained using a CCD (charge-coupled device) imager (Chemidock MP Imaging, Bio-Rad). The results presented in Fig. 3b are representative of a total of five western blots on independent biological samples. Immunolocalization of AOX was generally done as described in ref. 34. Briefly, cells were fixed with 2% formaldehyde in culture media for 20 min, washed three times with marine phosphate buffer (mPBS, see ref. 42) and permeabilized by 1% Triton X-100 in mPBS for 10 min. The cells were washed again, blocked for 30 min in 1% BSA in mPBS, and incubated overnight at room temperature (25°C) with anti-AOX antibody from rabbit (custom design, Sdix, 1:200 dilution in mPBS). The cells were then rinsed with mPBS and incubated with donkey Alexa 488-conjugated anti-rabbit IgG antibody (Life Technologies, at 1:100 dilution in mPBS) for 2 h at room temperature. Cells were then stained with $0.5 \mu\text{g ml}^{-1}$ DAPI (4',6'-diamidino-2-phenylindole, Life Technologies) for 10 min and mounted with Vectashield (Vector Laboratories) after a rinse. Finally the cells were observed using a Leica SP5 confocal microscope (Leica Microsystems).

Electron microscopy. For transmission electron microscopy, *P. tricornutum* cells were fixed in 0.1 M cacodylate buffer (Sigma-Aldrich), pH 7.4, containing 2.5% glutaraldehyde (TAAB), 2% formaldehyde (Polysciences) for 1 h at room temperature and then prepared according to a modified protocol from T. J. Deerinck *et al.* (<http://ncmir.ucsd.edu/sbfsem-protocol.pdf>). After the dehydration steps, the cells were infiltrated with ethanol/Epon resin mixture (2/3–1/3

for 1 h and 1/3–2/3 for 1 h) and finally embedded in Epon in a 60°C oven for 48 h or longer. Ultrathin sections (60 nm) were prepared with a diamond knife on an UC6 Leica ultramicrotome and collected on 200 μm mesh nickel grids before examining on a JEOL 1200 EX electron microscope.

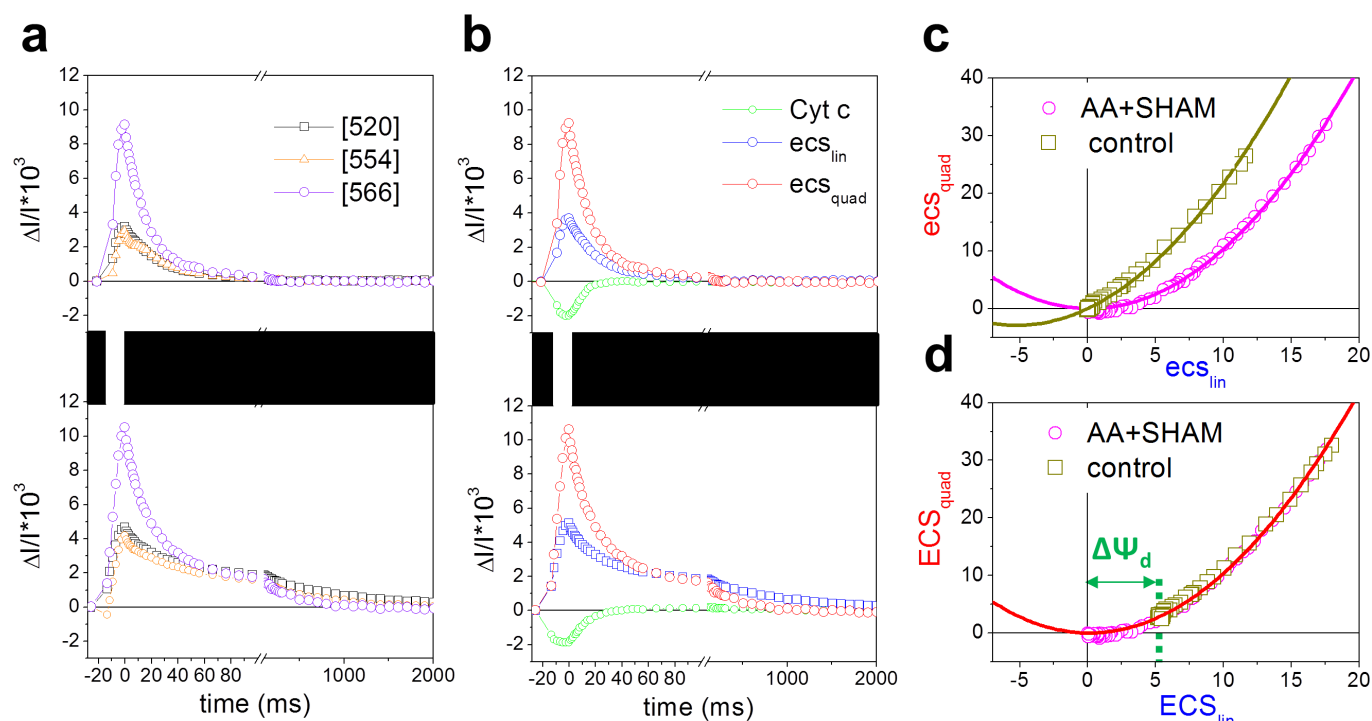
AOX knockdown lines. Partial coding sequence of AOX (identity Phatr2_bd1075) was amplified using primers AOXcFwd (TGCTCCGGAGGACAATGAATTCGC) and AOXcRev (TGGTCTAGACGTCGCGATGTTTC), cut by EcoRI/XbaI (Fermentas) and cloned into an EcoRI/XbaI-digested antisense construct⁴³, under control of the FcpB promoter. The AOX antisense construct was introduced into wild-type *P. tricornutum* cells using the standard microparticle bombardment procedure with a PDS-1000/He Particle Delivery System (Bio-Rad)^{43,44}. After 2 days, cells were transferred to F/2-supplemented-filtered seawater plates containing 100 $\mu\text{g ml}^{-1}$ phleomycin, a selective antibiotic for the *Sh ble* gene (InvivoGen, 09K30-MM). Putative antisense lines were verified for the presence of the transformed DNA by PCR screening using the primers ShbleFwd (ACCACTGCCGTTCCGGTG) and ShbleRev (TCGGTCAGTCTGCTCCTC), and the most strongly silenced lines were chosen on the basis of reduction of AOX protein levels as measured by SDS-PAGE western blot with an anti-AOX antibody (see section on western blots and immunolocalization).

25. Vartanian, M., Descl  s, J., Quinet, M., Douady, S. & Lopez, P. J. Plasticity and robustness of pattern formation in the model diatom *Phaeodactylum tricornutum*. *New Phytol.* **182**, 429–442 (2009).
26. Guillard, R. R. L. in *Culture of Marine Invertebrate Animals* (eds Smith W. L. & Chanley M. H.) 26–60 (Plenum, 1975).
27. Jo  t, T., Courmac, L., Horvath, E. M., Medgyesy, P. & Peltier, G. Increased sensitivity of photosynthesis to antimycin A induced by inactivation of the chloroplast *ndhB* gene. Evidence for a participation of the NADH-dehydrogenase complex to cyclic electron flow around photosystem I. *Plant Physiol.* **125**, 1919–1929 (2001).
28. Wishnick, M. & Lane, M. D. Inhibition of ribulose diphosphate carboxylase by cyanide. Inactive ternary complex of enzyme, ribulose diphosphate, and cyanide. *J. Biol. Chem.* **244**, 55–59 (1969).
29. Nakano, Y. & Asada, K. Purification of ascorbate peroxidase in spinach chloroplasts; its inactivation in ascorbate-depleted medium and reactivation by monodehydroascorbate radical. *Plant Cell Physiol.* **28**, 131–140 (1987).
30. Asada, K., Takahashi, M. A. & Nagate, M. Assay and inhibitors of spinach superoxide dismutase. *Agric. Biol. Chem.* **38**, 471–473 (1974).
31. Joliot, P. & Delosme, R. Flash induced 529 nm absorption change in green algae. *Biochim. Biophys. Acta* **357**, 267–284 (1974).
32. Santabarbara, S., Redding, K. E. & Rappaport, F. Temperature dependence of the reduction of p-700⁺ by tightly bound plastocyanin *in vivo*. *Biochemistry* **48**, 10457–10466 (2009).
33. Melis, A. Kinetic analysis of P-700 photoconversion: effect of secondary electron donation and plastocyanin inhibition. *Arch. Biochem. Biophys.* **217**, 536–545 (1982).
34. Bailleul, B., Cardol, P., Breyton, C. & Finazzi, G. Electrochromism: a useful probe to study algal photosynthesis. *Photosynth. Res.* **106**, 179–189 (2010).
35. Johnson, X. *et al.* A new setup for *in vivo* fluorescence imaging of photosynthetic activity. *Photosynth. Res.* **102**, 85–93 (2009).
36. Genty, B., Briantais, J. M. & Baker, N. R. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. Acta* **990**, 87–92 (1989).
37. Bilger, W. & Bj  rkman, O. Role of the xanthophyll cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in leaves of *Hedera canariensis*. *Photosynth. Res.* **25**, 173–186 (1990).
38. Peltier, G. & Thibault, P. O_2 uptake in the light in *Chlamydomonas*. *Plant Physiol.* **79**, 225–230 (1985).
39. Kana, T. M. *et al.* A membrane inlet mass spectrometer for rapid and high-precision determination of N_2 , O_2 , and Ar in environmental water samples. *Anal. Chem.* **66**, 4166–4170 (1994).
40. Rivasseau, C. *et al.* Accumulation of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate in illuminated plant leaves at supraoptimal temperatures reveals a bottleneck of the prokaryotic methylerythritol 4-phosphate pathway of isoprenoid biosynthesis. *Plant Cell Environ.* **32**, 82–92 (2009).
41. Bligny, R. & Douce, R. NMR and plant metabolism. *Curr. Opin. Plant Biol.* **4**, 191–196 (2001).
42. Van de Meene, A. M. L. & Pickett-Heaps, J. D. Valve morphogenesis in the centric diatom *Rhizosolenia setigera* (Bacillariophyceae, Centrales) and its taxonomic implications. *Eur. J. Phycol.* **39**, 93–104 (2004).
43. De Riso, V. *et al.* Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37**, e96 (2009).
44. Falcitatore, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol. (NY)* **1**, 239–251 (1999).



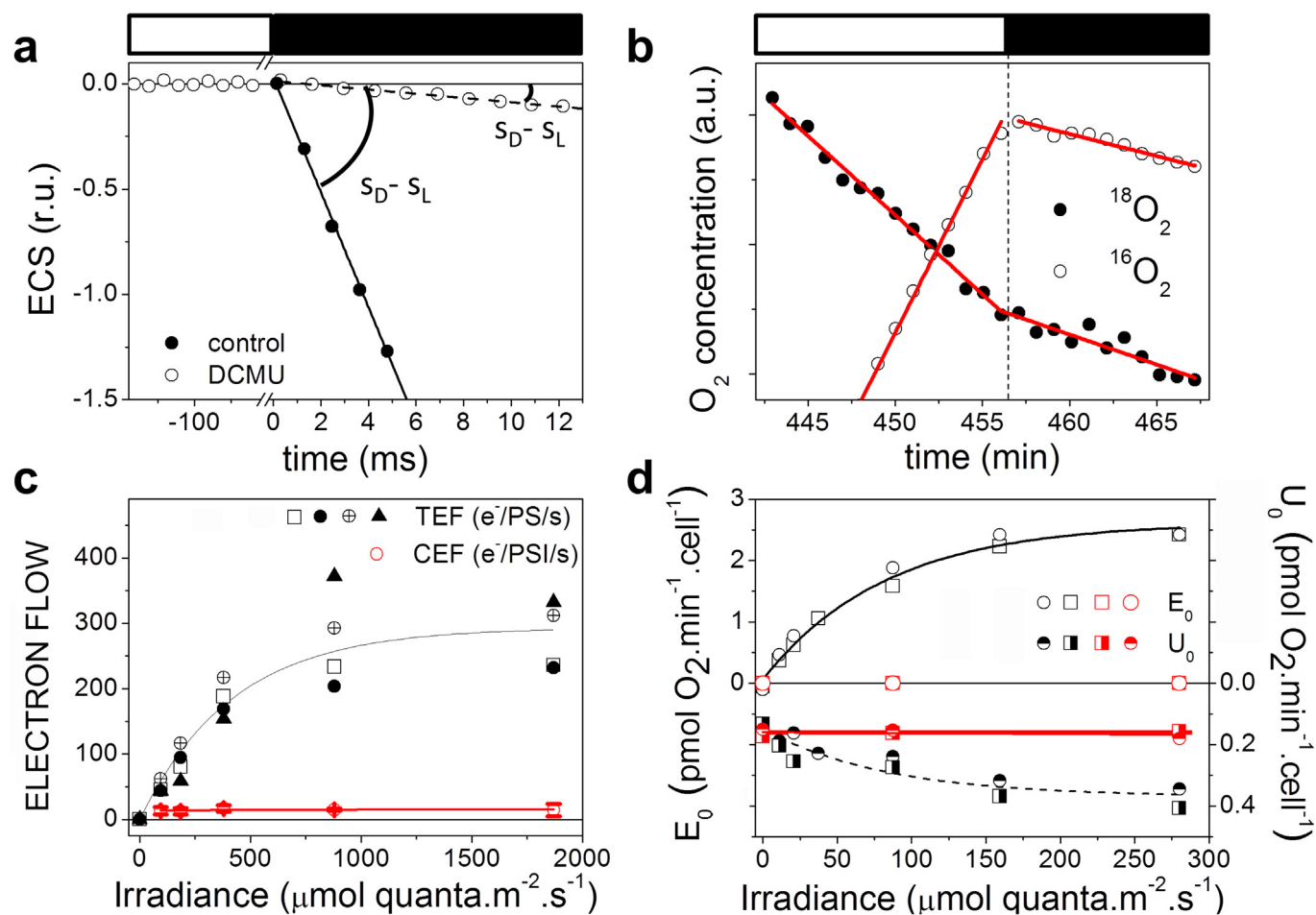
Extended Data Figure 1 | Deconvolution of the quadratic and linear ECS in *P. tricornutum*. **a**, Absorption difference ($\Delta I/I$) kinetics followed at different wavelengths in *P. tricornutum*, after a series of six saturating laser flashes, in anaerobic conditions. Solid lines correspond to the global fit of the experimental data with a sum of two exponential decays, with time constants τ and 2τ , respectively, as expected for linear and quadratic dependencies

(see Methods). **b**, $\Delta I/I$ spectra are shown at different times during ECS relaxation. All spectra were normalized to 1 at 520 nm for better comparison. The observation that the blue and green parts of the spectrum are homothetic during relaxation, while changes are seen in the red most part of it, reflects the presence of the two ECS components, having different relaxation kinetics.



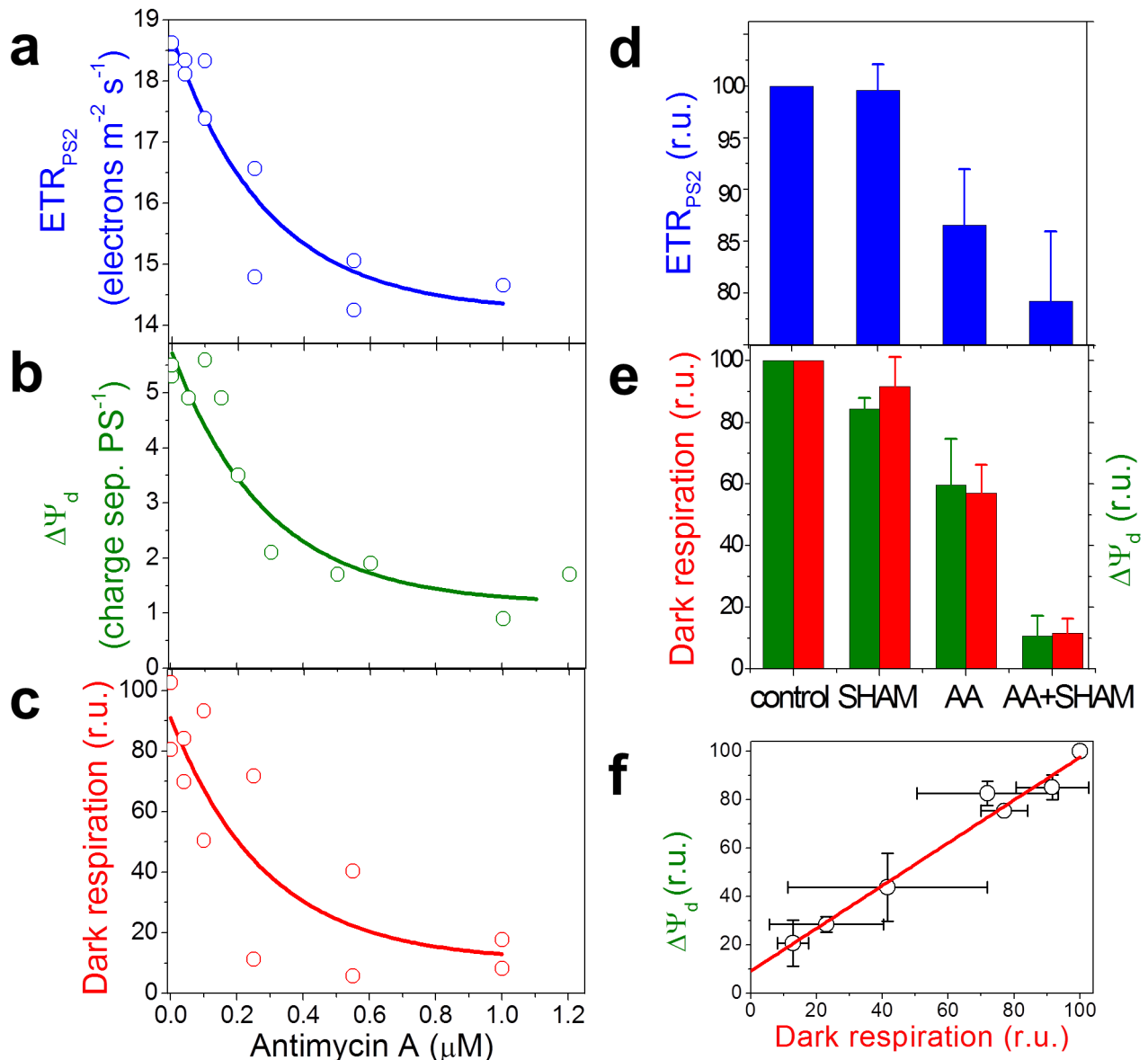
Extended Data Figure 2 | Separation of c-type cytochrome signals from linear and quadratic ECS signals in *P. tricornutum*. **a**, Kinetics of $\Delta I/I$ changes at 520, 554 and 566 nm during an ~10 ms pulse of saturating red light (4,500 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$) and the subsequent dark relaxation (top: control conditions; bottom: AA + SHAM). **b**, Kinetics of ecs_{lin} , ecs_{quad} changes and c-type cytochrome redox state, from kinetics in **a**, after deconvolution as explained in Methods. **c**, **d**, Relationship between the quadratic and the linear

ECS, before (ecs_{lin} , ecs_{quad} , **c**) and after (ECS_{lin} , ECS_{quad} , **d**) correction for the dark electric field (see Methods). Dark yellow and magenta symbols correspond to control and AA + SHAM conditions, respectively. The green arrow indicates the value of $\Delta\psi_d$ in control conditions. Data are representative of five independent biological samples. The black boxes in **a** and **b** indicate periods of darkness.



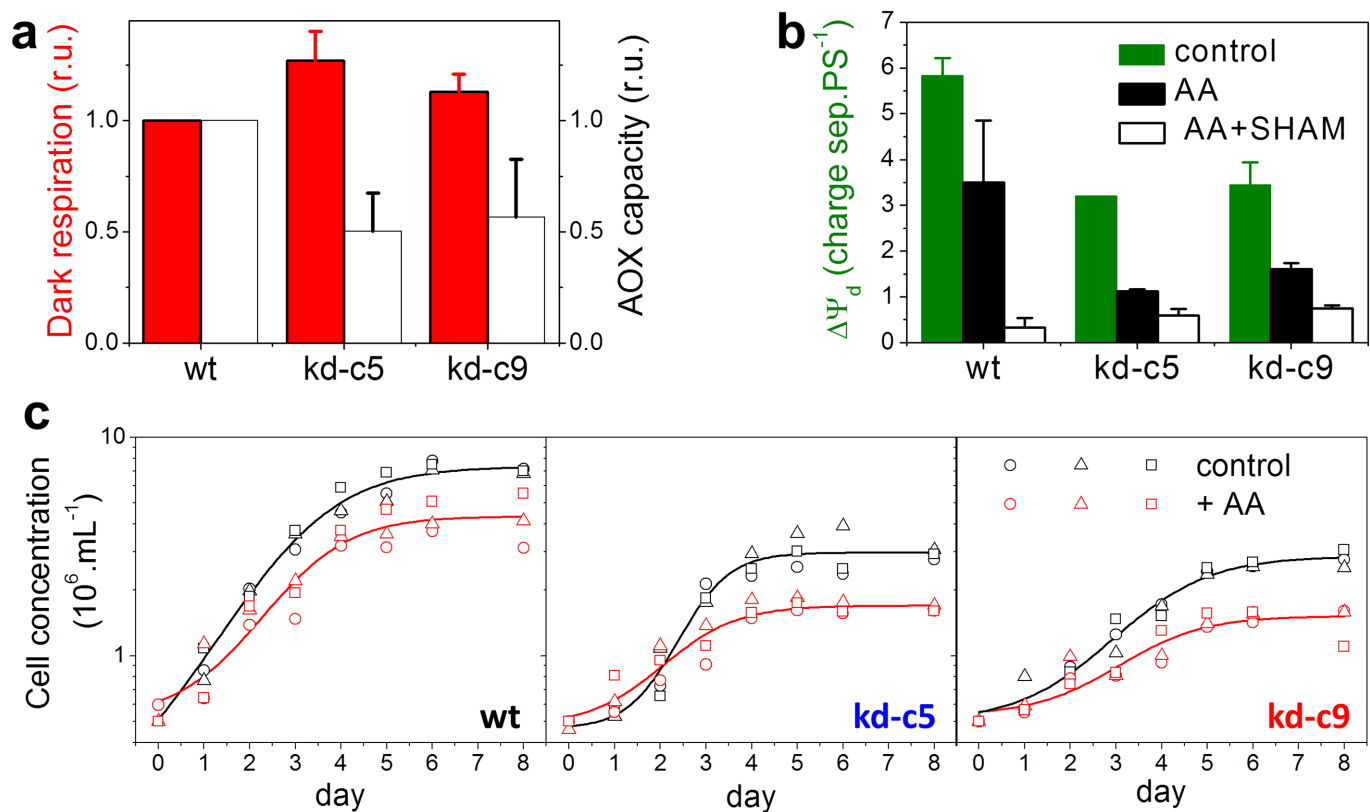
Extended Data Figure 3 | Cyclic electron flow and water-to-water cycles in *P. tricornutum*. **a**, Representative traces of changes in ECS_{lin} (normalized as explained in Methods, namely expressed in charge separation per PS) to evaluate linear and cyclic electron flow. Cells were illuminated with $1,870 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ of red light, in absence (filled circles) and presence (open circles) of DCMU and then transferred to the dark; r.u., relative units. **b**, Representative traces of the $^{16}\text{O}_2$ and $^{18}\text{O}_2$ concentrations at the offset of a $280 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ blue light; a.u., arbitrary units. In **a** and **b**, light and dark periods are represented by white and black boxes, respectively.

c, Photochemical rate corresponding to total electron flow (TEF, dark symbols, data from four independent biological samples) and CEF (red symbols, $n = 8 \pm \text{s.d.}$) at different irradiances. TEF and CEF were estimated from the initial slope of the ECS decay, as $(S_D - S_L)$ control and $(S_D - S_L)$ DCMU, respectively (see Methods). **d**, Light-dependencies of oxygen uptake (U_0 , half-filled symbols) and gross photosynthesis (E_0 , open symbols) in control conditions (dark) and in the presence of DCMU (red). Data from two independent biological samples (squares and circles).



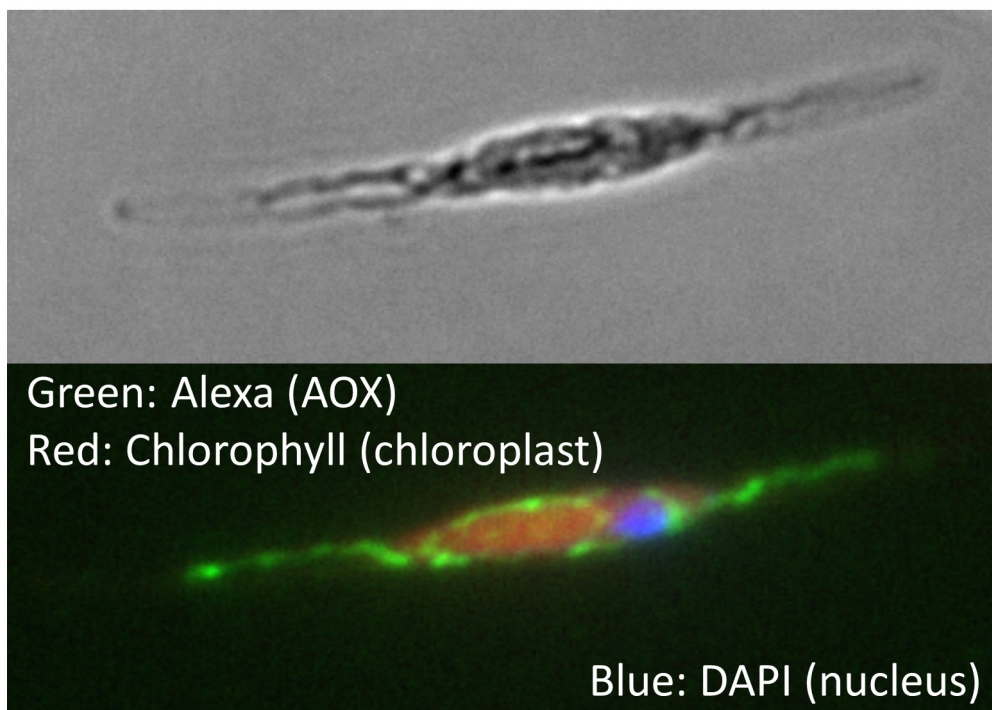
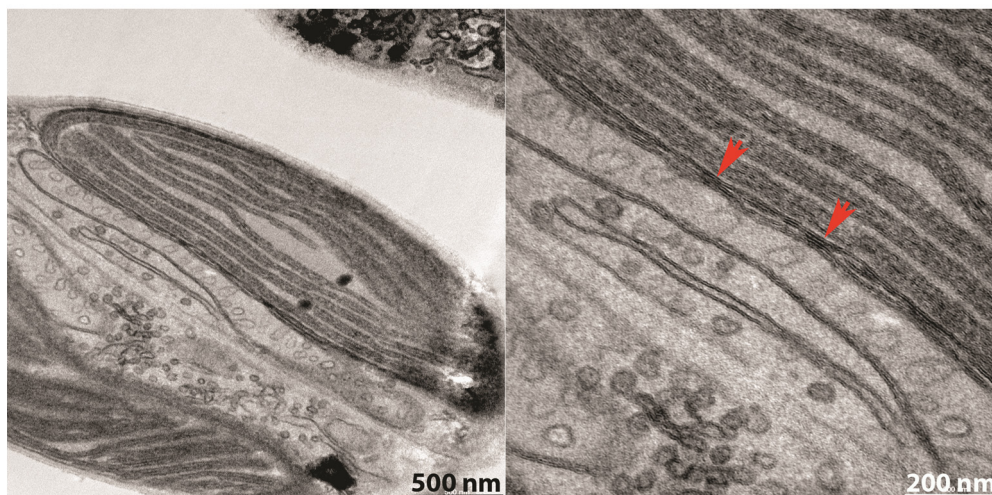
Extended Data Figure 4 | $\Delta\Psi_d$ and photosynthesis under respiratory inhibition in *P. tricornutum*. **a–c**, Dependency of the ETR_{PSII} (**a**), $\Delta\Psi_d$ (**b**) and dark respiration (**c**) upon inhibition of the cyanide-sensitive respiratory pathway with different concentrations of antimycin A, in the presence of saturating SHAM (1 mM). Data from two independent biological samples. Experimental data were fitted with a mono-exponential decay function.

d, e, ETR_{PSII} (**d**), $\Delta\Psi_d$ and dark respiration (**e**), expressed as percentage of control, in the presence of saturating AA (5 μM), SHAM (1 mM) or AA + SHAM (four independent experiments \pm s.d.). **f**, Relationship between $\Delta\Psi_d$ and mitochondrial respiration in samples treated with increasing concentrations of AA in the presence of 1 mM SHAM (mean value \pm s.e.m. from **b** and **c**).



Extended Data Figure 5 | Dark respiration, PMF and growth in AOX knockdown lines of *P. tricornutum*. **a**, Respiratory activity of wild-type and AOX knockdown lines. Total respiration rate (red bars) and the contribution of the AOX capacity (white bars, see Methods) were normalized to wild-type values ($n = 5 \pm \text{s.d.}$). **b**, ECS-based measurements of $\Delta\Psi_d$ in wild-type ($n = 3 \pm \text{s.d.}$) and AOX knockdown lines ($n = 2 \pm \text{s.e.m.}$), in control

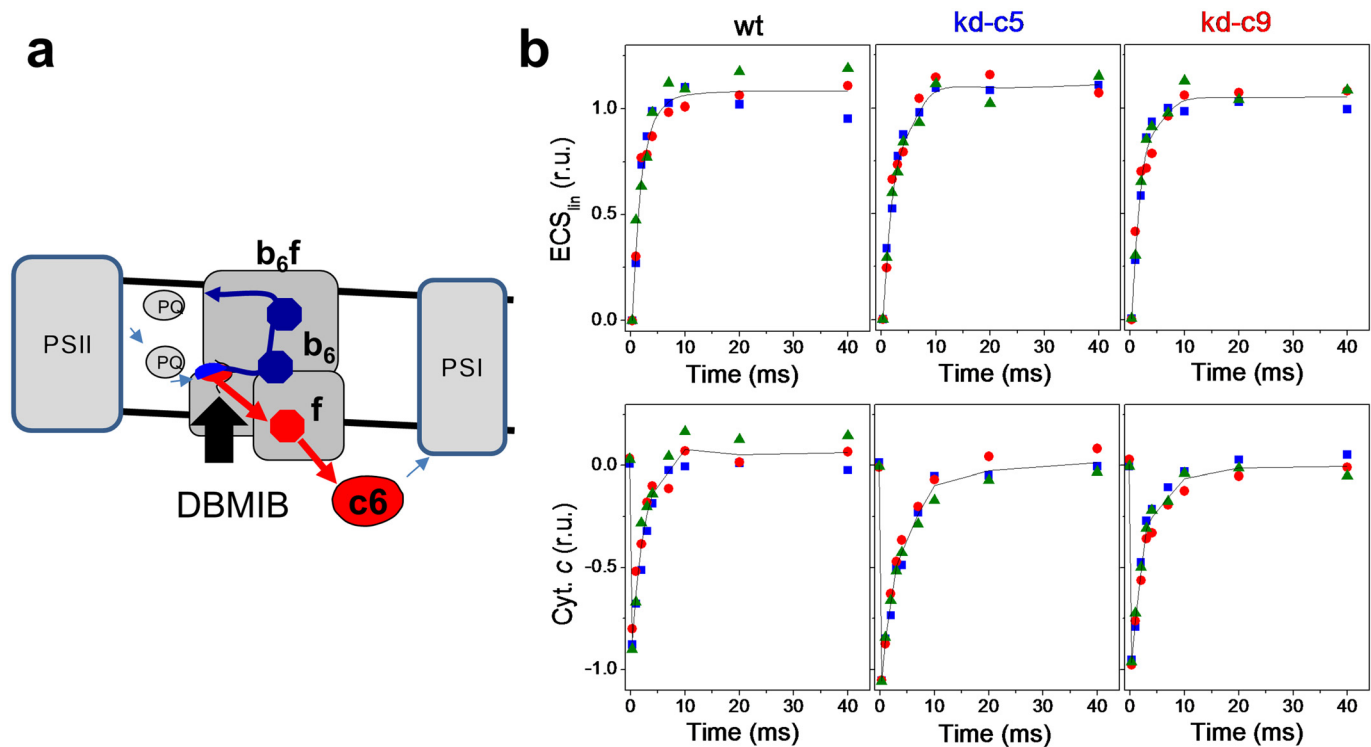
conditions (green), in the presence of AA (dark green) and in the presence of AA + SHAM (white). **c**, Growth curves of wild-type and AOX knockdown lines in the presence/absence of AA ($2 \mu\text{M}$). Three independent growth curves are shown for each strain/condition. AA was added every day and cells were grown in continuous light to prevent them from dying in the dark because of lack of respiration.

a**b**

Extended Data Figure 6 | Subcellular localization of AOX in *P. tricornutum* and plastid-mitochondria interaction in *P. tricornutum* wild-type cells.

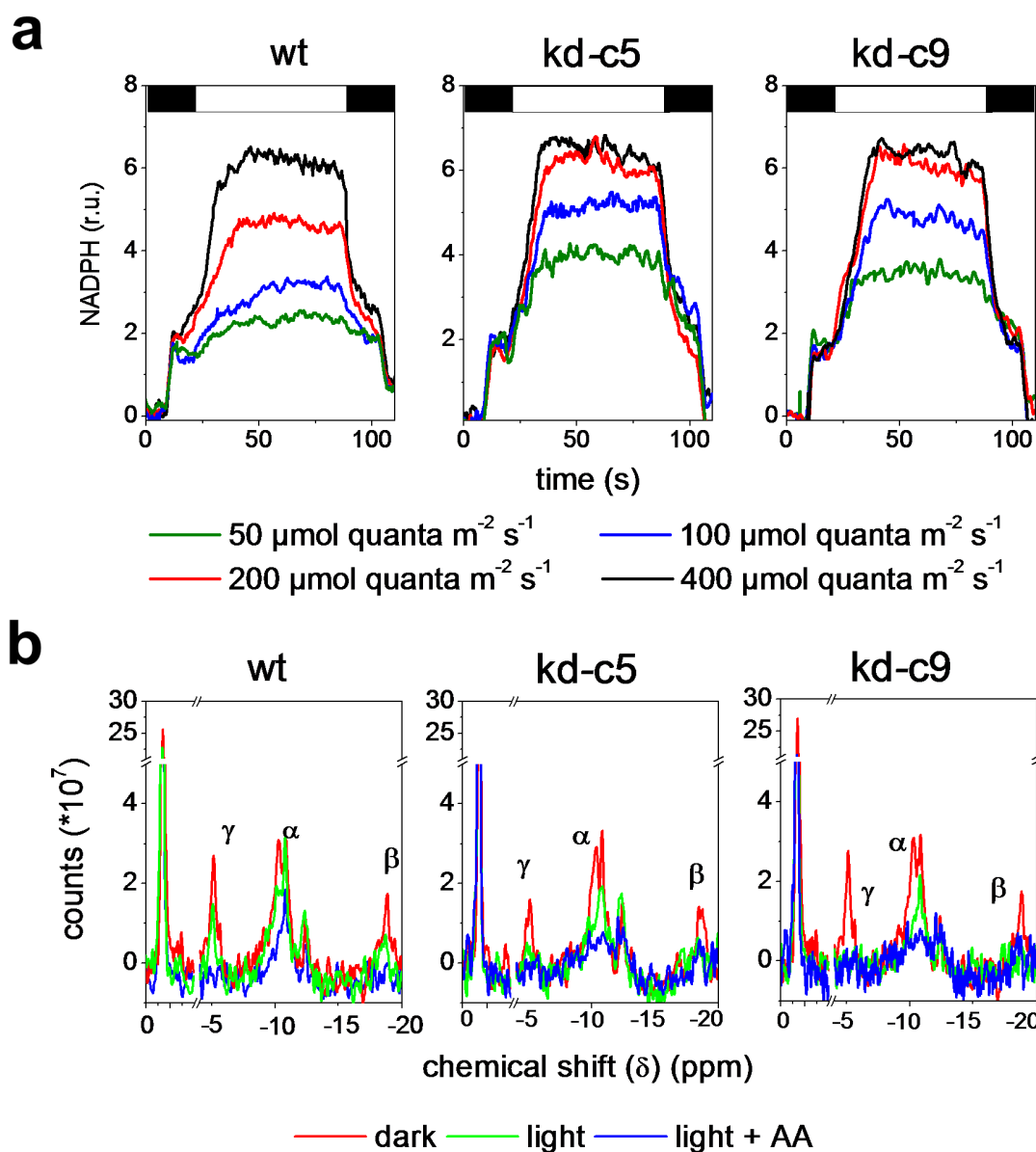
a, Subcellular localization of AOX. Cells were treated with an anti-AOX antibody and then with a secondary Alexa Fluor 488 antibody (see Methods). Positions of plastid and nuclei are indicated by chlorophyll *a* autofluorescence (red) and DAPI staining (blue), respectively. The pattern of AOX

localization is highly similar to that observed with the mitochondria-specific mito-tracker probe (data not shown). Images are representative of 60 cells from three independent biological samples. **b**, Electron micrographs of the plastid-mitochondria juxtaposition in *P. tricornutum*. Arrows indicate possible physical contacts between the plastid and mitochondrial membranes. Image is representative of 51 images from seven independent biological samples.



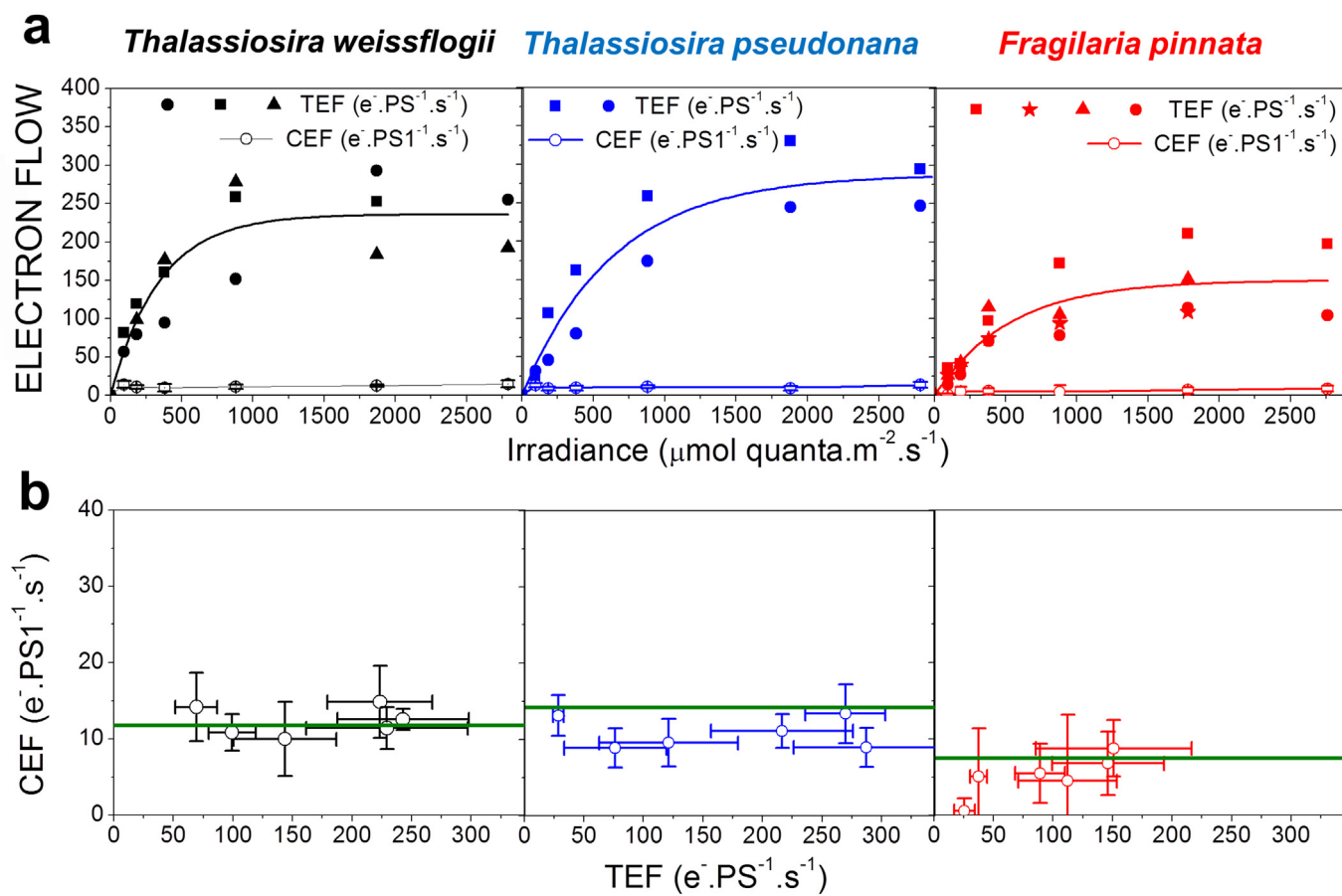
Extended Data Figure 7 | Cytochrome b_6f turnover in *P. tricornutum* wild-type and AOX knockdown lines. **a**, Schematic representation of the electron-flow reaction steps in the cytochrome b_6f complex, which can be evaluated by spectroscopic measurements. **b**, Slow phase of ECS_{lin} indicating cytochrome b_6 activity (top) and time-resolved redox changes of cytochromes c_6/f (bottom) in wild-type and AOX knockdown lines (kd-c5 and kd-c9). *P. tricornutum* cells were exposed to saturating single-turnover laser flashes

given 10 s apart. Data were normalized to the amplitude of the fast phase of the ECS_{lin} signal. Cytochrome c and ECS_{lin} were deconvoluted as explained in Methods. Three independent biological samples are shown in red, blue and green colours. Cell concentration was 2×10^7 cells per millilitre. Note that both the slow phase of the ECS_{lin} and the reduction of cytochromes c_6/f were completely abolished by the plastoquinone competitive inhibitor DBMIB (10 μ M; black arrow).



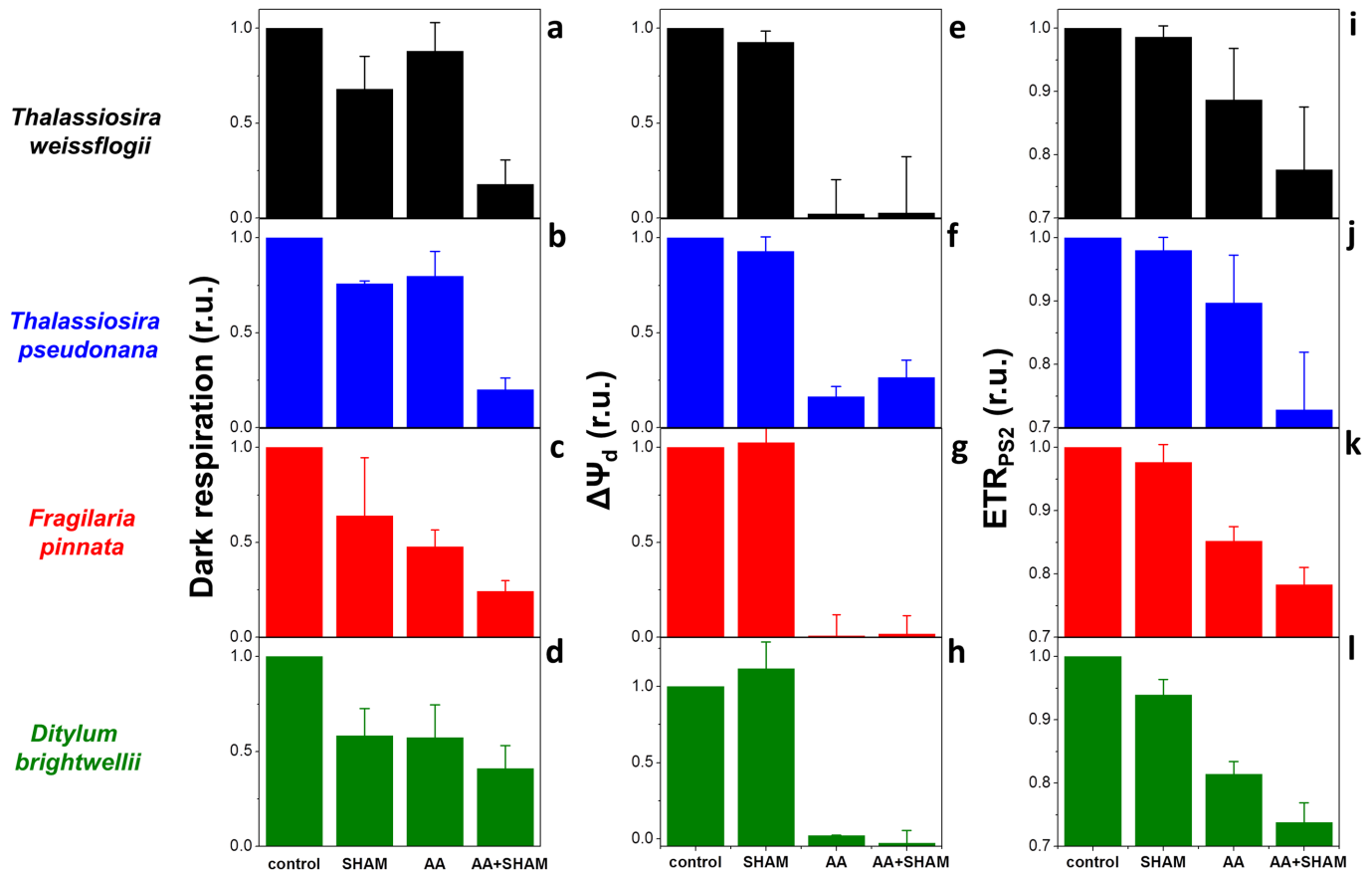
Extended Data Figure 8 | *In vivo* changes in the NADPH redox state and ATP in wild-type and AOX knockdown lines. **a**, Changes in NADPH at different light intensities. Light and dark periods are represented by white and black boxes, respectively. Light intensities were 50, 100, 200 and 400 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ (green, blue, red, and black traces, respectively). Chlorophyll concentration was $\sim 5 \mu\text{g ml}^{-1}$. **b**, Spectra from cells of wild-type (left) and

AOX knockdown lines c5 (middle) and c9 (right) in the dark (red), light (green) and light + AA (blue) conditions are shown, with normalization to the internal standard (methylenediphosphonate; pH 8.9). The positions of the α -, β - and γ -phosphates of NTPs are shown. **a**, **b**, Representative of three independent biological samples.



Extended Data Figure 9 | Cyclic electron flow in representative diatoms.
a, Total electron flow (filled symbols) was measured at different light irradiances, as illustrated in Extended Data Fig. 3, in *T. weissflogii* (black, $n = 3$ independent biological samples), *T. pseudonana* (blue, $n = 2$) and *F. pinnata*

(red, $n = 4$). Cyclic electron flow capacity was also measured for every species (open circles, five independent biological samples \pm s.d.). **b**, CEF capacity was plotted against TEF. CEF and TEF are presented as mean values \pm s.d. from **a**. The green line corresponds to CEF = 5% of the maximal total electron flow.



Extended Data Figure 10 | $\Delta\Psi_d$ and photosynthesis under conditions of respiratory inhibition in representative diatoms. Dark respiration (a–d), $\Delta\Psi_d$ (e–h) and ETR_{PSII} (i–l) in untreated cells and after treatment with inhibitors of respiration antimycin A, and/or SHAM at saturating concentrations. a, e, i, *T. weissflogii* (black). b, f, j, *T. pseudonana* (blue).

c, g, k, *F. pinnata* (red). d, h, l, *D. brightwellii* (green). The data represent the mean value \pm s.d. of three (l), four (d), five (f, h, k), six (a, c, e, j), seven (b, i) or eight (g) independent experiments. All data were normalized to the control value.

Mechanism of phospho-ubiquitin-induced PARKIN activation

Tobias Wauer¹, Michal Simicek¹, Alexander Schubert¹ & David Komander¹

The E3 ubiquitin ligase PARKIN (encoded by *PARK2*) and the protein kinase PINK1 (encoded by *PARK6*) are mutated in autosomal-recessive juvenile Parkinsonism (AR-JP) and work together in the disposal of damaged mitochondria by mitophagy^{1–3}. PINK1 is stabilized on the outside of depolarized mitochondria and phosphorylates polyubiquitin^{4–8} as well as the PARKIN ubiquitin-like (Ubl) domain^{9,10}. These phosphorylation events lead to PARKIN recruitment to mitochondria, and activation by an unknown allosteric mechanism^{4–12}. Here we present the crystal structure of *Pediculus humanus* PARKIN in complex with Ser65-phosphorylated ubiquitin (phosphoUb), revealing the molecular basis for PARKIN recruitment and activation. The phosphoUb binding site on PARKIN comprises a conserved phosphate pocket and harbours residues mutated in patients with AR-JP. PhosphoUb binding leads to straightening of a helix in the RING1 domain, and the resulting conformational changes release the Ubl domain from the PARKIN core; this activates PARKIN. Moreover, phosphoUb-mediated Ubl release enhances Ubl phosphorylation by PINK1, leading to conformational changes within the Ubl domain and stabilization of an open, active conformation of PARKIN. We redefine the role of the Ubl domain not only as an inhibitory¹³ but also as an activating element that is restrained in inactive PARKIN and released by phosphoUb. Our work opens up new avenues to identify small-molecule PARKIN activators.

The RING-between-RING E3 ligase PARKIN contains a RING1 domain that binds ubiquitin-charged E2 enzymes, and transfers ubiquitin from the E2 to an active site Cys residue in the RING2 domain and subsequently to a substrate. Cytosolic PARKIN exists in an autoinhibited, ‘closed’ conformation^{13–16}, in which binding to E2 is blocked by the amino-terminal Ubl domain as well as by a ‘repressor’ element (REP), and access to the RING2 active site Cys is blocked by the unique PARKIN domain (UPD, also known as RING0) (Extended Data Fig. 1). PhosphoUb binding and/or PARKIN Ubl phosphorylation are presumed to induce conformational domain rearrangements to activate PARKIN^{13–16}; however, the mechanism and sequence of events are unclear. Once activated, PARKIN ubiquitinates numerous mitochondrial and cytosolic proteins¹⁷, eventually triggering mitophagy.

To understand how phosphoUb induces PARKIN activation, we used PINK1-phosphorylated ‘ubiquitin suicide probes’¹⁸ that can modify Cys residues near a ubiquitin binding site *in vitro*¹⁹ (Fig. 1a and Extended Data Fig. 1d). Probes could not modify a previously crystallized construct of human PARKIN lacking the Ubl domain¹⁴ (HsPARKIN(ΔUbl), amino acids 137–465) (Fig. 1b). Unexpectedly, a similar fragment of *Pediculus humanus corporis* (human body louse) PARKIN (amino acids 140–461, hereafter referred to as *Ph*PARKIN) was modified by a subset of phosphoUb (pUb) suicide probes (Fig. 1b), enabling purification of the *Ph*PARKIN–pUb complex and determination of a crystal structure at 2.6 Å resolution (Fig. 1c, Methods and

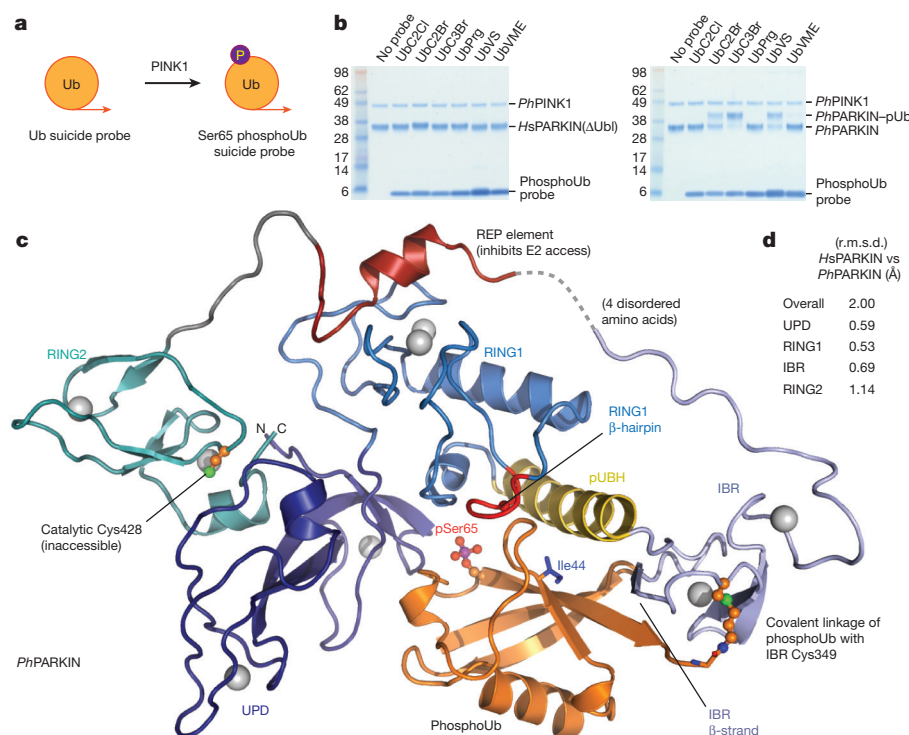


Figure 1 | Generation and structure of *Ph*PARKIN–phosphoUb complex. **a**, Schematic of the generation of phosphoUb suicide probes used in this study. **b**, HsPARKIN(ΔUbl) (left, amino acids 137–465) and *Ph*PARKIN (right, amino acids 140–461) were incubated with indicated ubiquitin suicide probes¹⁸ for 1 h in the presence of *Ph*PINK1 (Methods and Extended Data Fig. 1d) and resolved on Coomassie-stained SDS–PAGE gels. UbC2Cl, Ub chloroethylamine; UbC2Br, Ub bromoethylamine; UbC3Br, Ub bromopropylamine; UbPrg, Ub propargyl; UbVS, Ub vinylmethylsulfone; UbVME, Ub vinylmethylester. The experiment was performed three times with consistent results. Molecular weight markers are in kDa. **c**, Structure of the *Ph*PARKIN–pUb complex with domains coloured from blue to cyan (UPD, RING1, IBR, RING2), grey zinc atoms, red REP, yellow phosphoUb binding helix (pUBH), and orange phosphoUb. The catalytic Cys in RING2, and key phosphoUb residues, are indicated. **d**, Root mean squared deviation (r.m.s.d.) values for *Ph*PARKIN in comparison to HsPARKIN (Protein Data Bank (PDB) accession 4BM9 (ref. 14)).

¹Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.

Extended Data Figs 1e and 2a). *Ph*PARKIN-pUb resembles autoinhibited structures of *Hs*PARKIN (Fig. 1c, d and Extended Data Fig. 2e), with key differences as described below. The phosphoUb suicide probe had modified Cys349 in the *Ph*PARKIN in-between-RING (IBR) domain (Figs 1c and 2a), which in *Hs*PARKIN corresponds to probe-unreactive Gln347. Notably, *Hs*PARKIN Q347C is modified by phosphoUb suicide probes (Fig. 2b), indicating a similar binding mode of phosphoUb in *Hs*PARKIN. Hence, our complex structure serves as model for phosphoUb binding to *Hs*PARKIN.

PhosphoUb forms an extended interface (1,150 Å², 25% of ubiquitin surface) with the RING1 and IBR domains in *Ph*PARKIN, and also

interacts with side chains of the UPD (Fig. 2a and Extended Data Fig. 3). Key interactions are formed (1) via the phosphate group, which is located in a pocket formed by His304, Arg307 and Tyr314 of *Ph*PARKIN (Fig. 2c); (2) via the hydrophobic Ile44 patch of phosphoUb, which binds to an extended helix in the RING1 domain (amino acids 311–329, hereafter referred to as phosphoUb binding helix (pUBH)); (3) via a conserved surface β -hairpin loop (amino acids 280–288) in RING1 that harbours AR-JP mutations; and (4) via the phosphoUb carboxy terminus, which forms an intermolecular parallel β -sheet with a β -strand of the IBR domain (Figs 1c and 2a). Most residues forming phosphoUb interactions in *Ph*PARKIN are

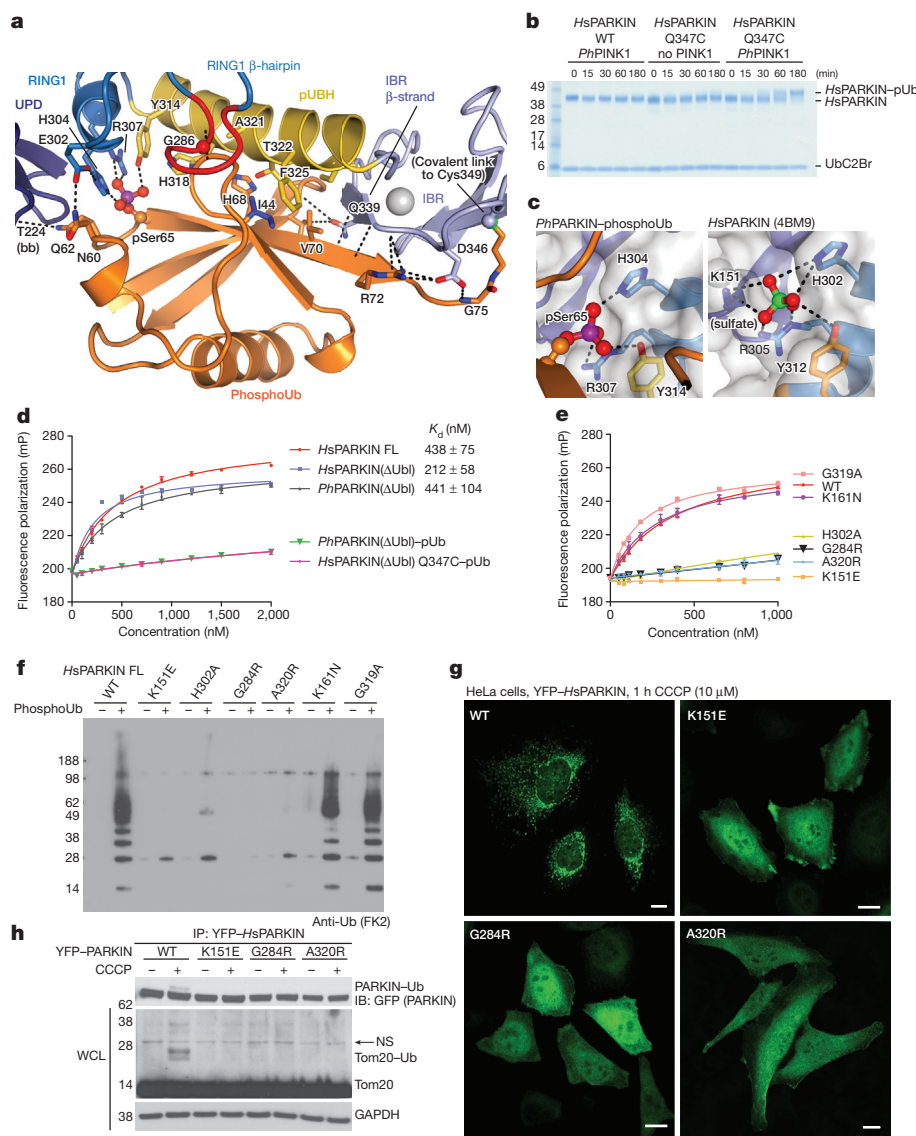


Figure 2 | PhosphoUb binding to PARKIN. **a**, PhosphoUb binding site on *Ph*PARKIN as in Fig. 1c. Dotted lines indicate hydrogen bonds. The RING1 β -hairpin that harbours patient mutations is highlighted in red. **bb**, backbone contacts. **b**, PhosphoUb suicide probe reactions as in Fig. 1b with UbC2Br and GST-*Ph*PINK1. The experiment was performed three times with consistent results. **c**, Occupied Ser65 phosphate pocket in *Ph*PARKIN (left); identical pocket in *Hs*PARKIN occupied by a sulfate ion in PDB 4BM9 (ref. 14) (right). **d**, Fluorescence polarization experiments characterizing the binding of FAsH-tagged phosphoUb to PARKIN variants. Measurements were performed in triplicate and error bars represent standard deviation from the mean. mP, millipolarization unit. **e**, Binding assays as in **d** with full-length *Hs*PARKIN and mutants in the phosphoUb binding site. Binding curves were compiled from experiments shown in Extended Data Fig. 3d. Measurements were performed in triplicate and error bars represent standard deviation from the mean.

f, Activity assays of full-length *Hs*PARKIN variants with and without phosphoUb. After 2 h, reactions were resolved by SDS-PAGE and polyubiquitin visualized by anti-polyubiquitin western blotting (FK2, Millipore). PARKIN protein normalization is shown in Extended Data Fig. 3e. The experiment was performed three times with consistent results. **g**, YFP-*Hs*PARKIN wild type or mutants were transfected into HeLa cells, treated with CCCP (10 μ M) for 1 h and visualized by immunofluorescence. See Extended Data Fig. 4 for controls and quantification. Scale bars, 10 μ m. **h**, HeLa cell lysates expressing YFP-*Hs*PARKIN wild type or mutants were western blotted for PARKIN (after immunoprecipitation (IP)) and Tom20 (in whole-cell lysate (WCL)). NS, non-specific band. The experiment was performed at least twice as biological replicate for every mutant with consistent results. See Extended Data Fig. 4d and Supplementary Information. Molecular weight markers are in kDa for **b**, **f**, **h**.

conserved in *HsPARKIN* (Extended Data Fig. 3a–c), and in our previous *HsPARKIN* structure¹⁴ the phosphate pocket is occupied by a sulfate molecule from the crystallization condition (Fig. 2c).

A fluorescence-polarization-based phosphoUb binding assay revealed sub-micromolar interactions of phosphoUb with full-length *Hs*PARKIN, *Hs*PARKIN(Δ Ubl) and *Ph*PARKIN (Fig. 2d). Modification with phosphoUb suicide probes of *Ph*PARKIN or *Hs*PARKIN Q347C (*Hs*PARKIN Q347C-pUb) abrogated phosphoUb binding (Fig. 2d), indicating that the covalently bound phosphoUb molecule satisfied the major phosphoUb binding site.

Mutations in the predicted phosphoUb interface in *HsPARKIN* reduced or abrogated phosphoUb binding (Fig. 2e and Extended Data Fig. 3d). *HsPARKIN* K151E (in the phosphate pocket; Fig. 2c, Ala152 in *PhPARKIN*), A320R (pUBH, Thr322 in *PhPARKIN*, Fig. 2a) or G284R (β -hairpin, Gly286 in *PhPARKIN*, Fig. 2a) abrogated phosphoUb binding (Fig. 2e). *HsPARKIN* G284R is an AR-JP-derived patient mutation, and our data provide a rationale for how this mutation leads to defects in PARKIN function (see below). Similarly, AR-JP mutation L283P (ref. 20) and cancer-associated H279P (ref. 21) in this region might also disrupt this loop and affect phosphoUb binding to *HsPARKIN*.

PARKIN activity can be assessed in autoubiquitination assays^{4-7,14}. *Hs*PARKIN was activated by phosphoUb, whereas *Hs*PARKIN K151E, H302A, A320R or G284R showed impaired phosphoUb-induced activation (Fig. 2f). *Hs*PARKIN K161N (an AR-JP mutation on the UPD¹⁴, see below) and *Hs*PARKIN G319A (see below) bound to and were activated by phosphoUb (Fig. 2e, f and Extended Data Fig. 3d, e).

In HeLa cells, carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP)-mediated depolarization of mitochondria led to rapid mitochondrial localization of yellow fluorescent protein (YFP)-tagged HsPARKIN, while phosphoUb-binding mutants did not show mitochondrial localization (Fig. 2g and Extended Data Fig. 4a–c). Moreover, wild-type HsPARKIN ubiquitinated endogenous Tom20 after CCCP treatment, while phosphoUb-binding mutants showed no apparent activity (Fig. 2h and Extended Data Fig. 4d).

Hence, we reveal that the phosphoUb binding site in PARKIN, which is conserved in divergent species, harbours AR-JP patient mutations and is important for PARKIN localization in cells (Fig. 2). This provides the molecular basis for PARKIN translocation to mitochondrial phosphoUb^{7,11,22} (Extended Data Fig. 3g).

Next we addressed the question of how phosphoUb activates PARKIN. PhosphoUb binds to a straight helix, pUBH, in *Ph*PARKIN (Figs 2 and 3). In previous PARKIN structures^{14–16} this helix is kinked at Gly319 (Ala321 in *Ph*PARKIN). The distinct conformation of the pUBH does not originate from this sequence difference; a crystal structure of *Hs*PARKIN(Δ Ubl) G319A still shows a kinked pUBH conformation (Extended Data Figs 1e and 5) and *Hs*PARKIN wild type and G319A have similar biochemical properties (Fig. 2e, f and Extended Data Fig. 3d, e). Notably, RING1 of the RBR E3 ligase HHARI (ref. 23) also features a kinked helix in the autoinhibited state (see Extended Data Fig. 5e–h). A kinked helix would be unable to bind phosphoUb, leading to a model in which the pUBH is dynamic and straightens upon phosphoUb binding. pUBH straightening hardly affects RING1 (Extended Data Fig. 6a) but impacts on the position of the IBR domain (Figs 2 and 3a, b), which rotates and moves by >20 Å as compared to full-length rat PARKIN (*Rn*PARKIN)¹⁶ (Fig. 3a). The conformational change stretches the IBR–REP linker (14–15 amino acids, Extended Data Fig. 2a) from spanning 31 Å in *Rn*PARKIN to cover a distance of 43 Å in *Ph*PARKIN–pUb (Fig. 3b and Extended Data Fig. 6b). This appears to destabilize inhibitory interactions of REP and RING2 as suggested by increased B-factors for these domains in *Ph*PARKIN–pUb (Extended Data Fig. 6c, d).

More importantly, phosphoUb binding also destabilizes the interface between the PARKIN Ubl domain and the RBR core, due to displacement of the IBR domain and reorganization of the IBR-REP linker that no longer spans the Ubl surface (Fig. 3b). Using isothermal titration calorimetry (ITC), we detected binding of isolated *Hs*PARKIN Ubl (amino acids 1–72) to *Hs*PARKIN(Δ Ubl) in *trans* (Fig. 3c, dissociation constant (K_d) ~ 40 μ M). Notably, binding is undetectable in the

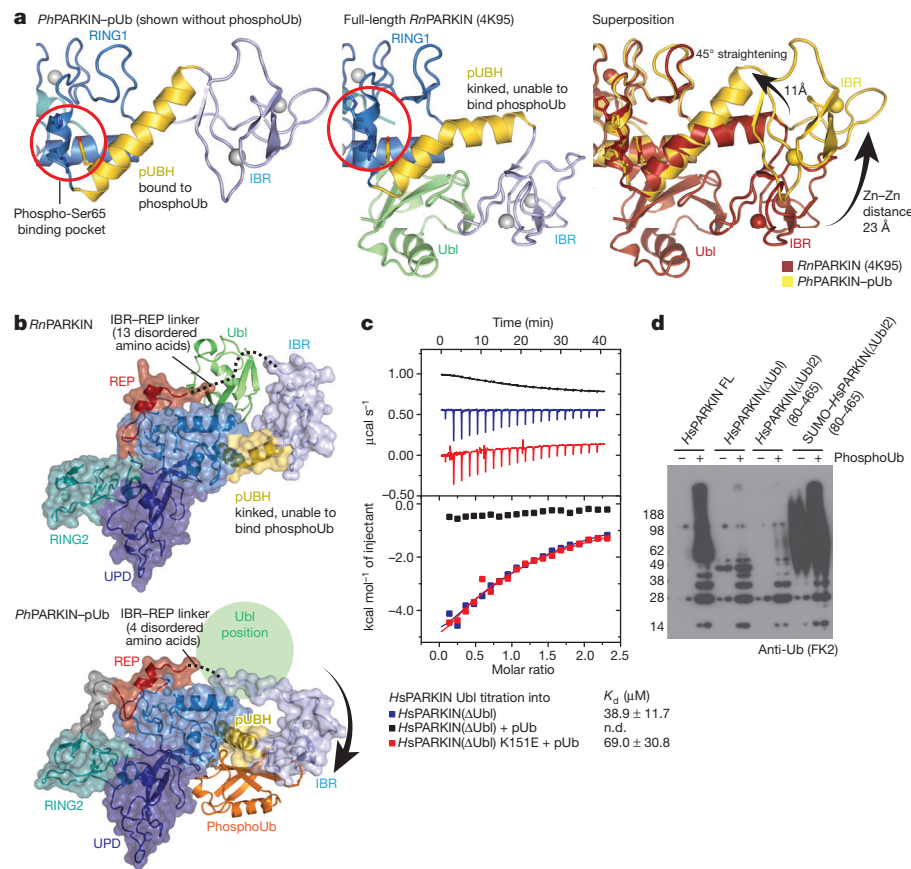


Figure 3 | Conformational changes due to phosphoUb binding. **a**, Straight pUBH in *Ph*PARKIN–pUb (left) versus kinked pUBH in previous PARKIN structures (middle; shown is full-length *Rn*PARKIN, PDB 4K95 (ref. 16)). Superposition indicating IBR repositioning by $>20 \text{ \AA}$ (right). **b**, Full-length *Rn*PARKIN (top) with core domains under a transparent surface and Ubl as green cartoon, and *Ph*PARKIN–pUb (bottom), with a green circle indicating the putative Ubl binding site; phosphoUb is in orange. **c**, ITC experiment titrating *Hs*PARKIN Ubl domain into *Hs*PARKIN(Δ Ubl) without phosphoUb (blue curve, $K_d \sim 39 \text{ }\mu\text{M}$), with phosphoUb present (black curve, K_d not detectable (n.d.)), and *Hs*PARKIN K151E with phosphoUb present (red curve, $K_d \sim 69 \text{ }\mu\text{M}$). Heat release curves (top panel) were shifted in their y -axis position for better visibility. Dissociation constant and curve fitting errors are indicated. All measurements were performed three times with consistent results. **d**, Activity of full-length *Hs*PARKIN, *Hs*PARKIN(Δ Ubl) (137–465), *Hs*PARKIN(Δ Ubl2) (80–465) and SUMO-tagged *Hs*PARKIN(Δ Ubl2) (80–465) with or without phosphoUb for 1 h (see Fig. 2f and Extended Data Fig. 3f). Experiments were performed three times with consistent results. Molecular weight markers are in kDa.

presence of phosphoUb, but recovered with phosphoUb-binding-deficient *HsPARKIN*(Δ Ubl) K151E mutant (Fig. 3c). Hence, phosphoUb binding releases the Ubl from the PARKIN RBR core.

PARKIN variants lacking the Ubl domain are still autoinhibited^{14–16} and less well activated by phosphoUb in comparison to full-length PARKIN⁷ (Fig. 3d). This indicates that the presence of the Ubl domain is important for full PARKIN activity. Interestingly, replacing the Ubl domain (amino acids 1–79) with SUMO, which lacks a Ubl-like hydrophobic patch and would not bind RING1, activates PARKIN constitutively, even in the absence of phosphoUb. This suggests that the released Ubl domain actively helps to unravel the autoinhibited PARKIN conformation. This, together with destabilization of the REP and RING2 autoinhibitory interactions, enables RING1 to bind and discharge E2–Ub conjugates⁴, and explains how PARKIN is activated by phosphoUb.

An alternative mechanism to activate PARKIN is PINK1-mediated phosphorylation of Ser65 in the Ubl domain^{7,9,10}, which was also sug-

gested to release the Ubl domain from the PARKIN core^{9,24}. In the closed conformation of PARKIN, the Ubl domain binds via its Ile44 patch to RING1 (refs 13, 16) (Fig. 4a and Extended Data Fig. 1). Importantly, PINK1 did not phosphorylate *HsPARKIN* Ubl I44A (or ubiquitin I44A) efficiently (Extended Data Fig. 7a, b). This suggests that PARKIN and PINK1 utilize overlapping binding sites on the Ubl domain, and that the Ubl domain has to be released from the PARKIN core for PINK1 to access and phosphorylate it.

Consistently, using a phosphospecific antibody against PARKIN phospho-Ser65 (anti-pSer65 PARKIN), we found that PARKIN phosphorylation is significantly enhanced when phosphoUb is added to the reaction, and this depends on ubiquitin phosphorylation and phosphoUb binding (Fig. 4c and Extended Data Fig. 7c, d). Induced release of the Ubl domain by mutating the binding site on RING1 (*HsPARKIN* L266R, Fig. 4a) leads to phosphorylation by PINK1 in the absence of phosphoUb (Fig. 4c and Extended Data Fig. 7c, d),

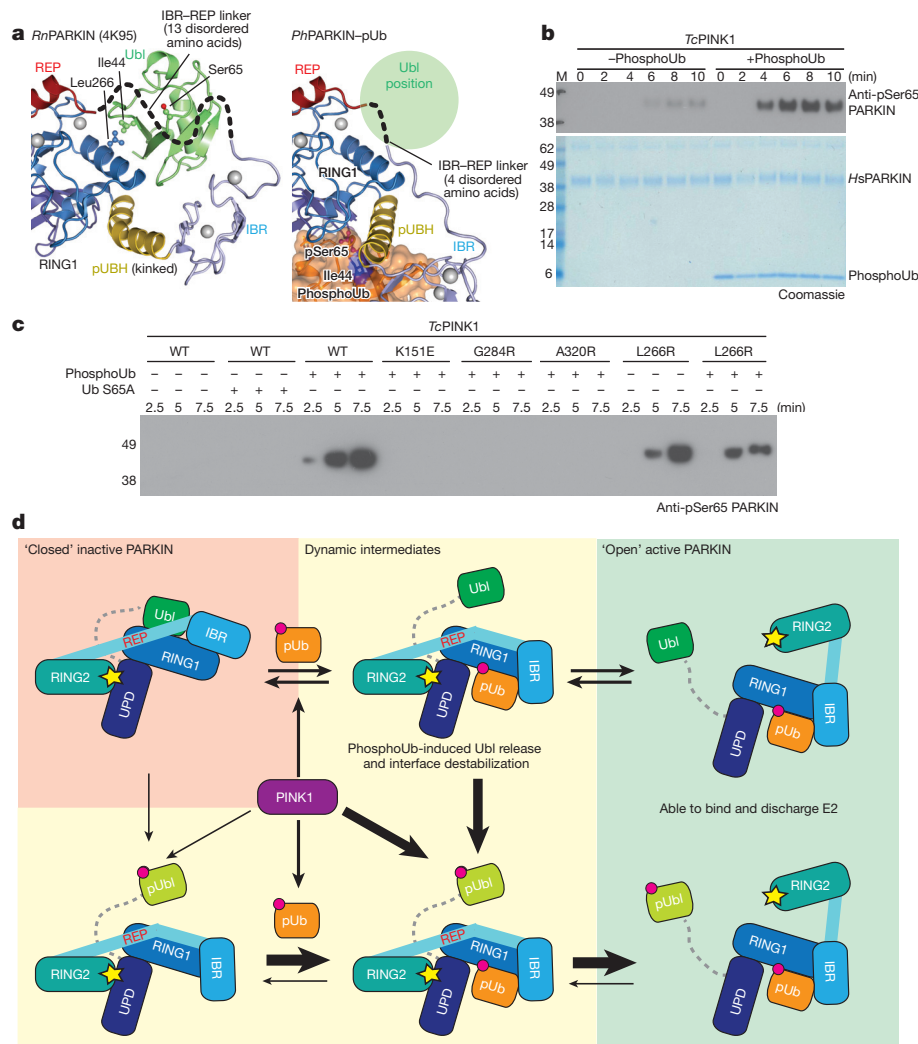


Figure 4 | PhosphoUb induces PARKIN Ubl phosphorylation and model of PARKIN activation. **a**, Detail of the Ubl position on PARKIN. Ubl Ile44 and interacting Leu266 on RING1 are highlighted. The mostly disordered IBR-REP linker hovers above Ser65 and moves out of the way in *PhPARKIN*-pUb (compare Fig. 3b). **b**, Time course of phosphorylation of *HsPARKIN* by GST-TcPINK1 (*Tribolium castaneum* PINK1), in the absence and presence of phosphoUb at 37 °C, visualized with anti-pSer65 PARKIN antibody (Abcam catalogue number ab154995) (top) and Coomassie (bottom). The experiment was performed three times with consistent results. **c**, Experiment as in **b** but performed at 22 °C with lower phosphoUb concentration (5.8 μ M). The experiment was performed three times with consistent results. Also see Extended Data Fig. 7c, d. **d**, Model of PARKIN activation. PARKIN

is autoinhibited (top left) by multiple mechanisms (Extended Data Fig. 1). Inactive PARKIN can interact with phosphoUb, which releases the Ubl domain, destabilizes inhibitory interactions and 'opens' PARKIN (top row). This process is reversible. PARKIN can also be phosphorylated by PINK1, either directly (bottom left) or with improved kinetics after phosphoUb binding (bottom middle). Phosphorylated Ubl undergoes a conformational change (Extended Data Fig. 8), probably preventing it from reverting back to the autoinhibited state. Phosphorylation of PARKIN hence stabilizes an open, active conformation (bottom right). Thickness of arrows indicates preferred routes. Further information on the role of Ubl phosphorylation is in Extended Data Figs 9 and 10. Molecular weight changes are in kDa and c.

indicating that Ubl release is crucial for enhanced phosphorylation. These results reveal a new function for phosphoUbl, namely to enable phosphorylation of the Ubl domain (Fig. 4). However, PINK1 can phosphorylate PARKIN in the absence of phosphoUbl *in vitro* (albeit inefficiently, Fig. 4b) and in cells^{7,25}, showing that PARKIN is a dynamic molecule (Fig. 4d) in which the Ubl domain is partially accessible by PINK1.

We next examined consequences of Ubl phosphorylation in HsPARKIN. Nuclear magnetic resonance measurements reveal that the Ubl domain undergoes significant changes when phosphorylated, in particular in the Ser65 loop and the Ile44 patch (Extended Data Fig. 8). Disruption of the Ile44 patch prevents re-binding of the Ubl domain to RING1^{13,24}, and probably stabilizes a more open, active conformation of PARKIN. Moreover, this may also explain why the phosphorylated Ubl domain cannot compete with phosphoUbl for the phosphoUbl binding site (Extended Data Fig. 9). Phosphorylated HsPARKIN does not impede binding of phosphoUbl, but rather enhances it⁷ (Extended Data Fig. 9).

Together, this completes our model of PARKIN activation (Fig. 4d). Autoinhibited PARKIN (Fig. 4d, top left) responds to phosphoUbl on mitochondria, which releases the Ubl to activate PARKIN in a reversible manner (Fig. 4d, top row). PINK1 phosphorylates PARKIN preferentially when the Ubl domain is released (Fig. 4d, bottom middle), and this leads to irreversible PARKIN activation (Fig. 4d, bottom right). Alternatively, inactive PARKIN may be phosphorylated by PINK1 directly^{7,25} (Fig. 4d, bottom left); this improves phosphoUbl binding, retains PARKIN on mitochondria, and irreversibly activates PARKIN (Fig. 4d, bottom row).

The conformation of fully active PARKIN remains elusive. Phosphorylated PARKIN but not phosphoUbl-activated PARKIN exposes its active site Cys residue⁷ (Extended Data Fig. 10), further indicating that PARKIN phosphorylation leads to ‘opening’ of PARKIN. Notably, mutations in a putative second phosphate pocket in the UPD, which we reported previously¹⁴ and which is distinct from the pocket involved in phosphoUbl binding, prevents phospho-Ubl-induced PARKIN opening and activation (Extended Data Fig. 10c, d). The functional link between PARKIN phosphorylation and a putative phosphate-binding pocket in the UPD may suggest that the phosphorylated Ubl domain binds back to the UPD; however, alternative activation mechanisms or the involvement of the Ubl-UPD linker cannot be excluded.

Our work is consistent with suggested models of PARKIN-mediated mitophagy^{7,25–28} and provides a structural understanding for phosphoUbl binding and allosteric PARKIN activation. We refine the role of the PARKIN Ubl domain as an essential activating element that is restrained in autoinhibited PARKIN. The model ensures tight temporal and spatial regulation of PARKIN activity, and incorporates a commitment step whereby PARKIN phosphorylation locks PARKIN in the active, open conformation. Our insights may prove useful pharmacologically, since small molecules that dislodge the PARKIN Ubl domain from the PARKIN core may activate PARKIN and benefit patients with Parkinson’s disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 February; accepted 3 July 2015.

Published online 10 July 2015.

1. Youle, R. J. & Narendra, D. P. Mechanisms of mitophagy. *Nature Rev. Mol. Cell Biol.* **12**, 9–14 (2011).
2. Corti, O., Lesage, S. & Brice, A. What genetics tells us about the causes and mechanisms of Parkinson’s disease. *Physiol. Rev.* **91**, 1161–1218 (2011).
3. Corti, O. & Brice, A. Mitochondrial quality control turns out to be the principal suspect in parkin and PINK1-related autosomal recessive Parkinson’s disease. *Curr. Opin. Neurobiol.* **23**, 100–108 (2013).
4. Koyano, F. *et al.* Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* **510**, 162–166 (2014).

5. Kane, L. A. *et al.* PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *J. Cell Biol.* **205**, 143–153 (2014).
6. Kazlauskaitė, A. *et al.* Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65. *Biochem. J.* **460**, 127–139 (2014).
7. Ordureau, A. *et al.* Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Mol. Cell* **56**, 360–375 (2014).
8. Wauer, T. *et al.* Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J.* **34**, 307–325 (2015).
9. Kondapalli, C. *et al.* PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating serine 65. *Open Biol.* **2**, 120080 (2012).
10. Shiba-Fukushima, K. *et al.* PINK1-mediated phosphorylation of the Parkin ubiquitin-like domain primes mitochondrial translocation of Parkin and regulates mitophagy. *Sci. Rep.* **2**, 1002 (2012).
11. Shiba-Fukushima, K. *et al.* Phosphorylation of mitochondrial polyubiquitin by PINK1 promotes Parkin mitochondrial tethering. *PLoS Genet.* **10**, e1004861 (2014).
12. Okatsu, K. *et al.* Phosphorylated ubiquitin chain is the genuine Parkin receptor. *J. Cell Biol.* **209**, 111–128 (2015).
13. Chaugule, V. K. *et al.* Autoregulation of Parkin activity through its ubiquitin-like domain. *EMBO J.* **30**, 2853–2867 (2011).
14. Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112 (2013).
15. Riley, B. E. *et al.* Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nature Commun.* **4**, 1982 (2013).
16. Trempe, J.-F. *et al.* Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science* **340**, 1451–1455 (2013).
17. Sarraf, S. A. *et al.* Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. *Nature* **496**, 372–376 (2013).
18. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
19. Wang, T. *et al.* Evidence for bidentate substrate binding as the basis for the K48 linkage specificity of otubain 1. *J. Mol. Biol.* **386**, 1011–1023 (2009).
20. Macedo, M. G. *et al.* Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson’s disease. *Mov. Disord.* **24**, 196–203 (2009).
21. Veeriah, S. *et al.* Somatic mutations of the Parkinson’s disease-associated gene *PARK2* in glioblastoma and other human malignancies. *Nature Genet.* **42**, 77–82 (2010).
22. Zheng, X. & Hunter, T. Parkin mitochondrial translocation is achieved through a novel catalytic activity coupled mechanism. *Cell Res.* **23**, 886–897 (2013).
23. Duda, D. M. *et al.* Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an ariadne-family E3 and insights into ligation mechanism. *Structure* **21**, 1030–1041 (2013).
24. Caulfield, T. R. *et al.* Phosphorylation by PINK1 releases the UBL domain and initializes the conformational opening of the E3 ubiquitin ligase Parkin. *PLOS Comput. Biol.* **10**, e1003935 (2014).
25. Ordureau, A. *et al.* Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. *Proc. Natl Acad. Sci. USA* **112**, 6637–6642 (2015).
26. Kazlauskaitė, A. & Muqit, M. M. K. PINK1 and Parkin — mitochondrial interplay between phosphorylation and ubiquitylation in Parkinson’s disease. *FEBS J.* **282**, 215–223 (2015).
27. Pickrell, A. M. & Youle, R. J. The roles of PINK1, Parkin, and mitochondrial fidelity in Parkinson’s disease. *Neuron* **85**, 257–273 (2015).
28. Koyano, F. & Matsuda, N. Molecular mechanisms underlying PINK1 and Parkin catalyzed ubiquitylation of substrates on damaged mitochondria. *Biochim. Biophys. Acta*. <http://dx.doi.org/10.1016/j.bbamer.2015.02.009> (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank M. Yu and beamline staff at Diamond Light Source, beamline I-04 and I-04-1, S. Freund and J. Wagstaff for NMR data, C. Johnson and S. McLaughlin for help with biophysics, BostonBiochem for providing UbVs and UbVME, C. Gladkova for help with cloning, and N. Birsa and J. Kittler (UCL London) for providing YFP-HsPARKIN plasmids. We thank members of the D.K. laboratory for reagents and discussions, and D. Barford, J. Pruneda and P. Elliott for comments on the manuscript. This work was supported by the Medical Research Council (U105192732), the European Research Council (309756), the Lister Institute for Preventive Medicine, the EMBO Young Investigator Program (to D.K.), and an EMBO Long-term Fellowship (to M.S.).

Author Contributions T.W. and D.K. designed the research, and T.W. performed all experiments. M.S. performed cell-based studies. A.S. contributed to characterization of Ubl and ubiquitin phosphorylation. T.W. and D.K. analysed the data and wrote the manuscript with help from all authors.

Author Information Coordinates and structure factors for the PhPARKIN-pUb complex and HsPARKIN(ΔUbl) G319A have been deposited with the Protein Data Bank under accession codes 5CAW and 5C9V, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.K. (dk@mrcc-imb.cam.ac.uk).

METHODS

No statistical methods were used to predetermine sample size.

Protein expression and purification. *HsPARKIN* full length, *HsPARKIN*(Δ Ubl) (amino acids 137–465), *PhPARKIN* (amino acids 140–461), *PhPINK1* (amino acids 115–575) as well as *Tribolium castaneum* PINK1 (*TcPINK1*, amino acids 128–570) were expressed as GST fusion proteins in Rosetta2 pLacI cells from pOPIN-K vectors as described¹⁴. In short, PARKIN cultures were induced by adding 200 μ M ZnCl₂ and 50 μ M IPTG, whereas PINK1 variants and the *HsPARKIN* Ubl domain (amino acids 1–72) were induced with 150 μ M IPTG followed by 12 h expression at 18 °C. After collection, the cell pellet was lysed by sonication in lysis buffer (270 mM sucrose, 10 mM glycerol 2-phosphate disodium, 50 mM NaF, 14 mM β -mercaptoethanol, 50 mM Tris (pH 8.0)), in the presence of lysozyme, DNase I and EDTA-free protease inhibitors. The suspension was centrifuged and the supernatant applied to Glutathione Sepharose 4B beads (GE Healthcare). After 1 h of agitation at 4 °C the beads were washed with high salt buffer (500 mM NaCl, 10 mM DTT, 25 mM Tris (pH 8.5)) and equilibrated in low salt buffer (200 mM NaCl, 10 mM DTT, 25 mM Tris (pH 8.5)). GST-tagged proteins were either eluted from beads with low salt buffer containing 40 mM glutathione and purified by gel filtration (Superdex 200, GE Life Sciences), or cleaved on beads by incubating with GST-3C protease for 12 h at 4 °C before further purification by gel filtration (Superdex 75, GE Life Sciences) in low salt buffer as a final step. When improved purity was required, such as for *PhPARKIN* used in crystallization, an anion exchange step (RESOURCE Q, GE Life Sciences) using a linear gradient of 75–600 mM NaCl in 10 mM DTT, 25 mM Tris (pH 8.5) was included before final gel filtration.

N-terminally His6-SUMO-tagged *HsPARKIN*(Δ Ubl2) (amino acids 80–465) was expressed as described above and lysed in His6-lysis buffer (200–300 mM NaCl, 4 mM β -mercaptoethanol, 25 mM Tris (pH 8.5)). After centrifugation, the protein was affinity purified with Talon Superflow resin (GE Healthcare) and eluted in 200 mM NaCl, 4 mM β -mercaptoethanol, 25 mM Tris (pH 8.5) containing 200–250 mM imidazole. Eluted protein was directly applied to gel filtration (Superdex 75, GE Life Sciences) in low salt buffer. Protein for SUMO-tag cleavage was dialysed overnight in 200 mM NaCl, 4 mM β -mercaptoethanol, 25 mM Tris (pH 8.5) containing His6-tagged SENP1. The sample was reappplied to Talon resin, and the flow-through purified by gel filtration (Superdex 75, GE Life Sciences) in low salt buffer.

Phospho-*HsPARKIN* for biochemical assays was generated by incubating 32–37 μ M *HsPARKIN*, 5.4 μ M GST-*PhPINK1* and 10 mM ATP with 1 \times ligation buffer (40 mM Tris pH 7.4, 10 mM MgCl₂, 0.6 mM DTT) for 1 h at room temperature. GST-*PhPINK1* was removed with Glutathione Sepharose 4B resin (GE Healthcare) and *HsPARKIN* was re-purified by gel filtration (Superdex 75, GE Life Sciences). Consistent phosphorylation levels were checked by western blot analysis using an anti-pSer65 PARKIN antibody (Abcam cat no. ab154995).

Modification with ubiquitin-based suicide probes. Probe reactions for biochemical assays were performed by incubating 5 μ M PARKIN with 40 μ M indicated ubiquitin suicide probe (Figs 1b and 2b) or 20 μ M Ub-VS (Extended Data Fig. 10), 1 μ M *PhPINK1* where indicated and 5 μ M phosphoUb where indicated in the presence of 1 \times reaction buffer (50 mM Tris pH 7.4, 200 mM NaCl, 10 mM DTT, 10 mM MgCl₂). The reaction took place at room temperature after adding 10 mM ATP and was quenched by adding LDS sample buffer at the indicated time points. Samples were applied on NuPAGE 4–12% Bis-Tris gels (Invitrogen) and stained with Instant Blue SafeStain (Expedeon).

To generate the covalent *PhPARKIN*-pUb complex for crystallization, 46 μ M *PhPARKIN* (amino acids 140–461) was reacted with 230 μ M UbC3Br probe and 23 μ M GST-*PhPINK1* in the presence of 1 \times reaction buffer. The coupling was initiated by adding 10 mM ATP and incubated for 6 h at room temperature. The complex was purified by gel filtration (Superdex 75, GE Life Sciences) in low salt buffer. Fractions containing *PhPARKIN*-pUb were pooled, concentrated and used for crystallization without freezing.

Crystallization, data collection and refinement. *PhPARKIN*-pUb was crystallized in sitting-drop vapour diffusion at a concentration of \sim 5.3 mg ml⁻¹ at 18 °C. Crystals were grown in 100 nl protein solution mixed with 100 nl mother liquor (2% (v/v) PEG400, 2 M ammonium sulfate, 0.1 M HEPES pH 7.5). Before vitrification, crystals were soaked in 1.7% (v/v) PEG400, 15% (v/v) glycerol, 1.7 M ammonium sulfate, 0.085 M HEPES (pH 7.5) for cryo-protection. Diffraction data were collected at the Diamond Light Source (Harwell, UK), beamline I-04, at 100 K and a wavelength of 0.97949 Å, to a resolution of 2.62 Å.

The *HsPARKIN*(Δ Ubl) (amino acids 137–465) G319A mutant was crystallized at a concentration of 2.4 mg ml⁻¹ by mixing 400 nl protein solution with 400 nl mother liquor (1.8 M lithium sulfate, 0.01 M MgCl₂, 0.05 M MES pH 5.6) in a sitting-drop vapour diffusion set-up at 18 °C. Before vitrification in liquid nitrogen the crystals were briefly soaked in 1.6 M lithium sulfate, 0.01 M MgCl₂, 0.05 M

MES pH 5.4 containing 15% (v/v) glycerol. Diffraction data were collected at the Diamond Light Source (Harwell, UK), beamline I04-1, at 100 K and a wavelength of 0.91730 Å, to a resolution of 2.35 Å.

Phasing of the *PhPARKIN*-pUb data set was performed by molecular replacement with Phaser²⁹ using isolated domains of *HsPARKIN*(Δ Ubl) (PDB 4BM9 (ref. 14)) and ubiquitin (PDB 1UBq (ref. 30)) as search models. The structure of *HsPARKIN*(Δ Ubl) G319A was solved by using *HsPARKIN*(Δ Ubl) (PDB 4BM9 (ref. 14)) as a refinement model. For both structures, subsequent rounds of model building in Coot³¹ and refinement in Phenix³² resulted in final models with statistics shown in Extended Data Fig. 1e. The *HsPARKIN*(Δ Ubl) G319A structure was refined with simulated annealing to reduce model bias. Both structures were refined with TLS, using different protein chains as individual TLS groups. Final Ramachandran statistics were 95.3%/4.7%/0.0% (favoured/allowed/outliers) for the *HsPARKIN* G319A mutant structure, and 96.7%/3.2%/0.1% for the *PhPARKIN*-pUb structure. Structure figures were generated with PyMol (<http://www.pymol.org>).

Fluorescence polarization phosphoUb binding assays. N-terminally FLAsH-tagged ubiquitin was phosphorylated and purified as described for phosphoUb⁸ with buffers supplemented with β -mercaptoethanol. Labelling was performed overnight at 4 °C with 60 μ M FLAsH-tagged phosphoUb at a ratio of 37.5:1 (v/v) with Lumio Green (Invitrogen) in 1 \times FLAsH dilution buffer (50 mM Tris (pH 7.4), 1 mM β -mercaptoethanol). Buffer exchange was performed with PD-10 desalting columns (GE Healthcare) in 50 mM Tris (pH 7.4), 2 mM β -mercaptoethanol and samples were concentrated to \sim 8 μ M FLAsH-labelled phosphoUb. For binding studies in 384-well low volume plates (Corning), 10 μ l of 100 nM labelled phosphoUb was mixed with 10 μ l of PARKIN serial dilutions in FLAsH buffer (20 mM Tris (pH 7.4), 100 mM NaCl, 2 mM β -mercaptoethanol, 0.1 mg ml⁻¹ bovine serum albumin). Fluorescence polarization (FP = $(I_{\parallel} - I_{\perp}) / (I_{\parallel} + I_{\perp})$) was measured using a PheraStar plate reader (BMB Labtech) with the optic module set to λ_{ex} = 485 nm and λ_{em} = 520 nm. Measurements were performed in triplicate and error bars are given as the standard deviation from the mean. A least square fit for one binding site was performed using the following equation

$$FP = (B_{\text{max}} \times X / (K_d + X)) + NS \times X + \text{background}$$

with FP being fluorescence polarization and X the concentration of the titrant, B_{max} is the maximum specific binding, K_d is the equilibrium dissociation constant and NS the slope for nonspecific binding, which was restricted to values greater than 0.

PARKIN activity assays. Spin-filtered *HsPARKIN* (2 μ M) was pre-incubated for 0.5 h at 30 °C with 10 mM ATP, 1 \times ligation buffer (40 mM Tris pH 7.5, 10 mM magnesium chloride, 0.6 mM dithiothreitol (DTT)), a total concentration of 0.5 mg ml⁻¹ ubiquitin and 0.05 mg ml⁻¹ phosphoUb or 0.1 μ M GST-*TcPINK1* where indicated. Ubiquitination was initiated by adding 0.1 μ M E1 and 1 μ M UBE2L3 (Fig. 2f and Extended Data Fig. 10d) or 0.2 μ M E1 and 2 μ M UBE2L3 (Fig. 3d). The reaction was quenched with LDS sample buffer containing DTT and iodoacetamide to prevent forming of disulfide bridges. NuPAGE 4–12% gradient Bis-Tris gels were used for separation, and proteins were transferred on a nitrocellulose membrane with subsequent detection using an anti-polyubiquitin FK2 antibody (Millipore).

Immunofluorescence. HeLa cells (originating from ATCC) were nucleofected with N-terminally tagged eYFP-*HsPARKIN* (gift from the J. Kittler laboratory) and grown on coverslips for 24–48 h. After treatment with DMSO or CCCP (10 μ M) for 1 h, cells were fixed with 4% paraformaldehyde, washed with 0.1 M glycine in phosphate buffered saline (PBS), at pH 7.4, briefly permeabilized with 0.2% Triton X-100 in PBS and blocked with a blocking solution containing 10% goat serum and 0.5% BSA. Samples were further incubated with anti-Tom20 antibody (FL-145, Santa Cruz) followed by goat Alexa647-coupled anti-rabbit antibody (Life Technologies). Confocal images were taken using a Zeiss LSM780 microscope.

Immunoblotting and immunoprecipitation. HeLa cells (originating from ATCC) were transfected with eYFP-*HsPARKIN*. After 24–48 h, cells were treated with DMSO or CCCP as before, and lysed in cell lysis buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1% (v/v) NP-40) supplemented with EDTA-free protease inhibitor cocktail (Roche), PhosphoSTOP (Roche), as well as 10 mM N-ethylmaleimide (Sigma-Aldrich) and 10 mM chloracetamide (Sigma-Aldrich) for better detection of ubiquitinated proteins. For immunoprecipitation of eYFP-*HsPARKIN*, 500 μ g of cell lysate was incubated with GFP-Trap agarose (Chromotek) for 1 h. The beads were washed three times with cell lysis buffer, and proteins were eluted with 1 \times LDS buffer. Cells were regularly checked for the absence of mycoplasma infection using the MycoAlert Kit (Lonza). Antibodies were from commercial sources: goat anti-GFP (ab6673, Abcam), rabbit anti-Tom20 (FL-145, Santa Cruz), mouse anti-GAPDH (6C5, Ambion).

Isothermal titration calorimetry. Isothermal titration calorimetry (ITC) experiments were performed in a MicroCal iTC200 machine (GE Healthcare) at 20 °C with the sample and the ligand in low salt buffer. The cell contained 35 μ M of

HsPARKIN(Δ Ubl) and 400 μ M *HsPARKIN* Ubl was injected in 2 μ l injections at 120 s intervals. Protein sample as well as ligand were in low salt buffer and *HsPARKIN*(Δ Ubl) was mixed with phosphoUb at a 1:1.2 molar ratio as indicated. Binding curves were integrated and fitted to a one-site binding model by using the MicroCal ITC Analysis plug-in for Origin (Malvern).

Phosphorylation assays. PARKIN phosphorylation was performed by incubating 5 μ M *HsPARKIN* with 0.5 μ M GST-TcPINK1, 10 mM ATP, 1 \times reaction buffer and phosphoUb (14 μ M unless stated differently). The reaction was quenched at the given time points with LDS sample buffer and proteins were separated on a NuPAGE 4–12% gradient Bis-Tris gel, transferred on nitrocellulose membrane and detected with anti-pSer65 PARKIN antibody (Abcam cat no. ab154995). Phosphorylation assays of *HsPARKIN* Ubl domain (amino acids 1–72) and ubiquitin were performed as described above with 20 μ M of *HsPARKIN* Ubl domain and ubiquitin, respectively. For the ubiquitin phosphorylation assay, the GST-TcPINK1 concentration was increased to 1.5 μ M. The reaction was quenched at the given time points with LDS sample buffer and proteins were separated on a 15% SuperSep Phos-tag gel (Wako Chemicals) and stained with Instant Blue SafeStain (Expedeon).

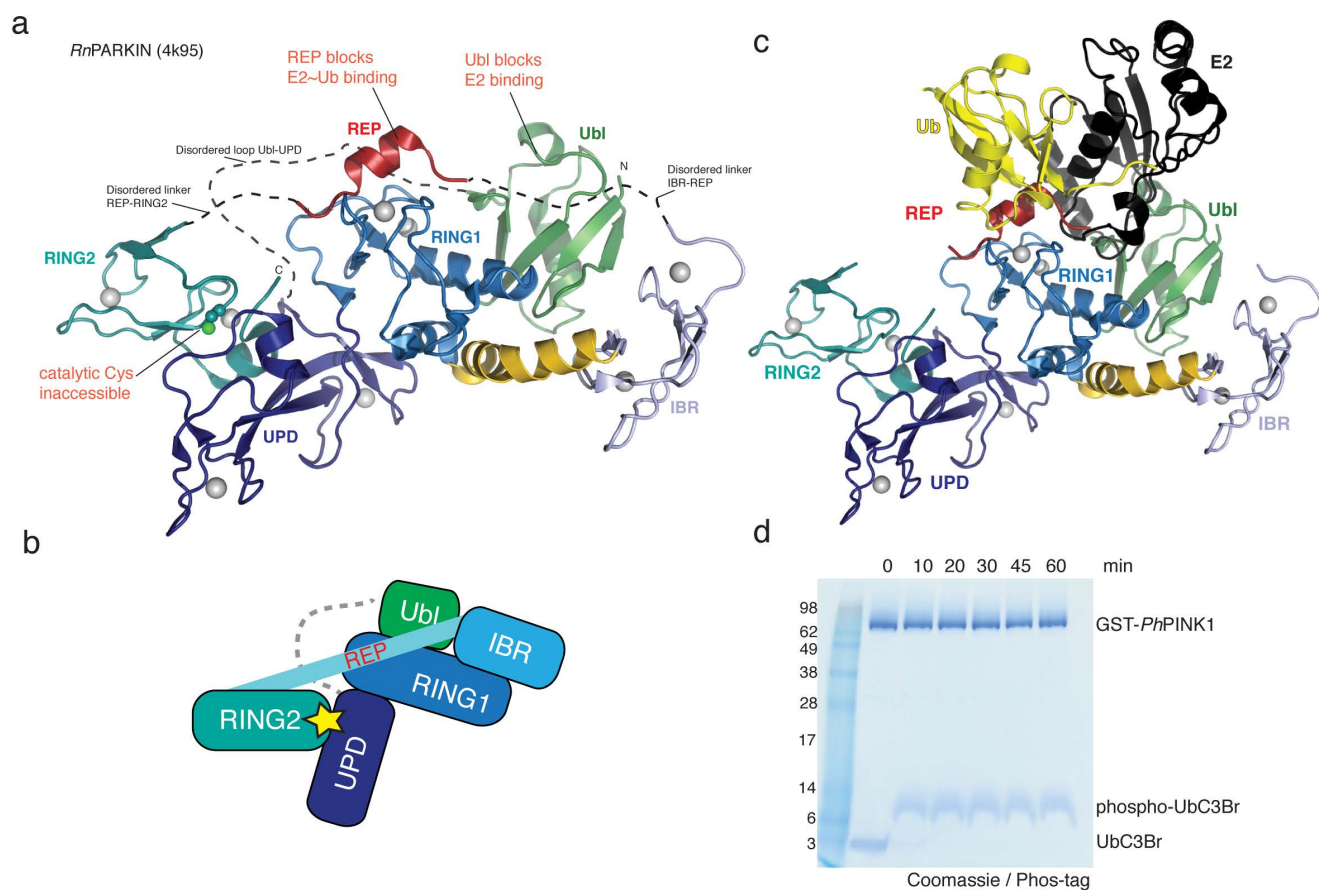
Phosphorylation of the PARKIN Ubl domain for NMR analysis. Isotope-labelled *HsPARKIN* Ubl domain (amino acids 1–72) was expressed in M9 minimal media supplemented with 4 g l⁻¹ ¹³C-glucose, 2 g l⁻¹ ¹⁵N-NH₄Cl, trace elements and BME vitamins (Sigma-Aldrich) and purified as described above. The final gel filtration was performed in NMR buffer (18 mM Na₂HPO₄, 7 mM NaH₂PO₄ (pH 7.2), 150 mM NaCl, 5 mM TCEP). Isotope-labelled *HsPARKIN* Ubl was phosphorylated at room temperature by incubating 80 μ M *HsPARKIN* Ubl with 2.5 μ M PhPINK1, 1 mM ATP and 1 \times ligation buffer which was adjusted to 332.5 μ l with NMR buffer, before addition of 17.5 μ l D₂O as lock solvent. The reaction was monitored by consecutive ¹H,¹⁵N 2D BEST-TROSY (band selective excitation short transients transverse relaxation optimized spectroscopy) experiments and quenched with apyrase.

Solution studies of the phosphorylated PARKIN Ubl domain. NMR acquisition was performed at 298 K on a Bruker Avance III 600 MHz spectrometer

equipped with a cryogenic triple resonance TCI probe. The software packages Topspin3.2 (Bruker) and Sparky (Goddard & Kneller, UCSF; <http://www.cgl.ucsf.edu/home/sparky/>) were used for data processing and analysis, respectively. ¹H,¹⁵N 2D BEST-TROSY experiments³³ were conducted with in-house optimized Bruker pulse sequences that contained a recycling delay of 400 ms and 512 \times 64 complex points in the ¹H,¹⁵N dimension, respectively.

Standard HSQC-based Bruker triple resonance pulse sequences were used to generate backbone chemical shift assignments. CBCACONH and HNCACB spectra were collected with 50% non-uniform sampling (NUS) of 1,024 \times 32 \times 55 complex points in the ¹H, ¹⁵N and ¹³C dimensions. HNCO and HNCACO experiments were acquired using NUS at a rate of 50% with 1,024 \times 32 \times 48 complex points in the ¹H, ¹⁵N and ¹³C dimensions, respectively. Data set processing was performed with compressed sensing using the MddNMR software package³⁴. Weighted chemical shift perturbation calculations were completed using the equation $\sqrt{((\Delta^1\text{H})^2 + (\Delta^{15}\text{N}/5)^2)}$.

29. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
30. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531–544 (1987).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
32. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
33. Favier, A. & Brutscher, B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR* **49**, 9–15 (2011).
34. Kazimierczuk, K. & Orekhov, V. Y. Accelerated NMR spectroscopy by using compressed sensing. *Angew. Chem. Int. Edn Engl.* **50**, 5556–5559 (2011).
35. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7–E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature Struct. Mol. Biol.* **19**, 876–883 (2012).
36. Sakata, E. *et al.* Parkin binds the Rpn10 subunit of 26S proteasomes through its ubiquitin-like domain. *EMBO Rep.* **4**, 301–306 (2003).
37. Kellsall, I. R. *et al.* TRIAD1 and HHARI bind to and are activated by distinct neddylated Cullin-RING ligase complexes. *EMBO J.* **32**, 2848–2860 (2013).



e

Data collection statistics. Numbers in brackets are for the highest resolution bin.

	<i>Ph</i> PARKIN~phosphoUb	<i>Hs</i> PARKIN G319A
Data collection		
Space group	<i>P</i> 2 ₁	<i>H</i> 3 ₂
<i>a</i> , <i>b</i> , <i>c</i> (Å)	49.17, 44.12, 204.53	169.36, 169.36, 96.99
α , β , γ (°)	90, 92.55, 90	90, 90, 120
Resolution (Å)	34.06 - 2.62 (2.74 - 2.62)	86.68 - 2.35 (2.43 - 2.35)
<i>R</i> _{merge}	0.085 (0.389)	0.089 (0.802)
$\langle I / \sigma I \rangle$	6.3 (2.0)	13.3 (2.1)
Completeness (%)	99.6 (99.6)	100 (100)
Redundancy	3.0 (3.1)	6.9 (6.8)
Refinement		
Resolution (Å)	34.06 - 2.62	84.68 - 2.35
No. reflections (test set)	26558 (1330)	22270 (1148)
<i>R</i> _{work} / <i>R</i> _{free}	0.228 / 0.260	0.198 / 0.229
No. atoms		
Protein	5569 (765 aa)	2357 (306 aa)
Ligand/ion	25	48
Water	33	81
<i>B</i> -factors		
Protein	70.2	54.3
Ligand/ion	115.1	88.9
Water	50.3	47.1
R.m.s deviations		
Bond lengths (Å)	0.005	0.003
Bond angles (°)	0.620	0.691

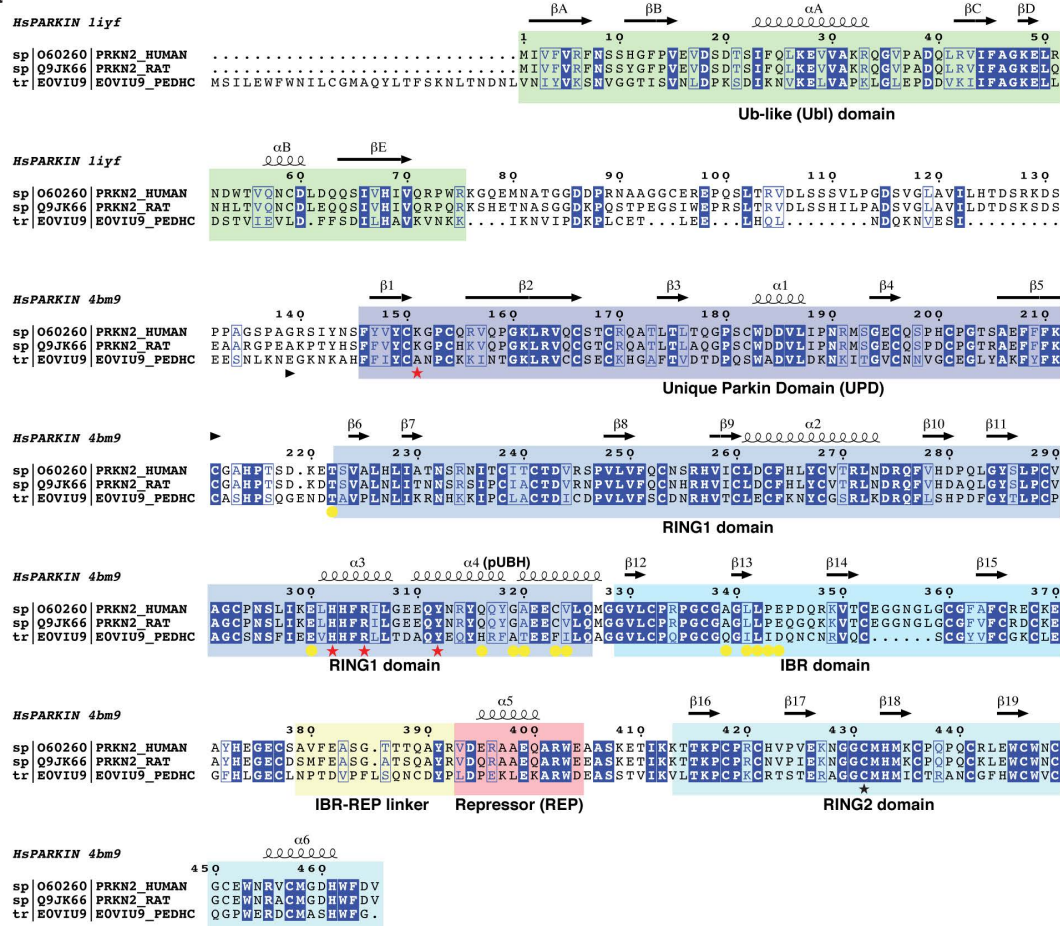
Datasets were collected and structures determined from a single crystal each.

Extended Data Figure 1 | Autoinhibited PARKIN and phosphoUb probes.

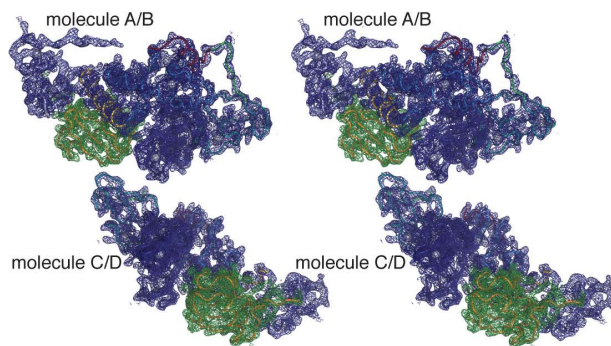
a, Structure of *Rn*PARKIN (PDB 4K95 (ref. 16), chain A is used for all representations of *Rn*PARKIN) with Ubl domain coloured in green, the unique PARKIN domain (UPD) in dark blue, RING1 in blue, IBR in light blue, REP in red, and RING2 in cyan. Zinc atoms are shown as grey spheres. The catalytic Cys431 is shown in ball-and-stick representation. Disordered linkers are indicated as dotted lines. The three mechanisms of PARKIN inhibition indicated in red. **b**, Schematic diagram of the closed, autoinhibited conformation of full-length PARKIN. **c**, Superposition of the E2–Ub complex from the BIRC7 structure (PDB 4AUQ (ref. 35)), superposed via its RING domain onto RING1 of full-length *Rn*PARKIN (PDB 4K95 (ref. 16)) to indicate

the position of the E2–Ub on PARKIN RING1. Assuming that E2–Ub adopts a canonical conformation on RING1, the E2 would clash with the Ubl domain and partially with the REP, while the E2-linked ubiquitin would clash with the REP. Hence, Ubl and REP have to be released to enable PARKIN E2–Ub binding at RING1. **d**, Time-course analysis of an exemplary reaction of UbC3Br (0.2 mg ml^{-1}) phosphorylated by GST–*Ph*PINK1 ($5 \mu\text{M}$) as described previously for ubiquitin⁸. Phosphorylation of proteins was monitored by a band shift on Coomassie-stained Phos-Tag gels. The experiment was performed two times with consistent results. Molecular weight markers are in kDa. **e**, Data collection and refinement statistics.

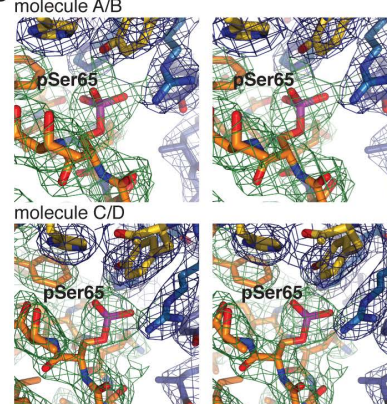
a



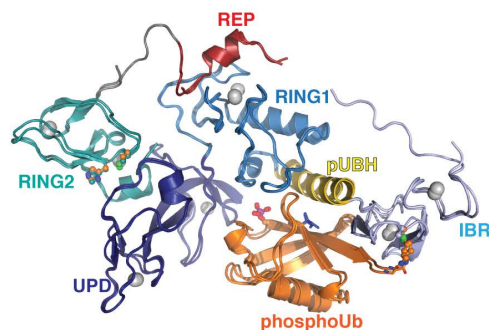
b



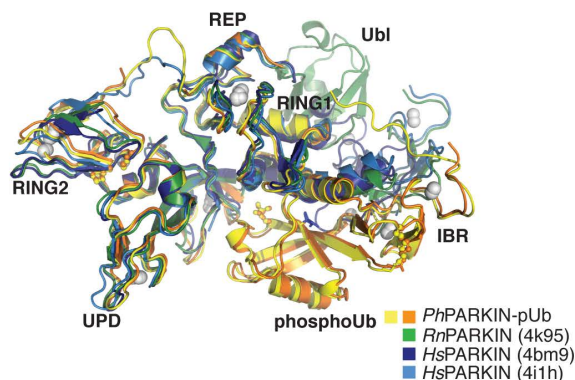
c



d

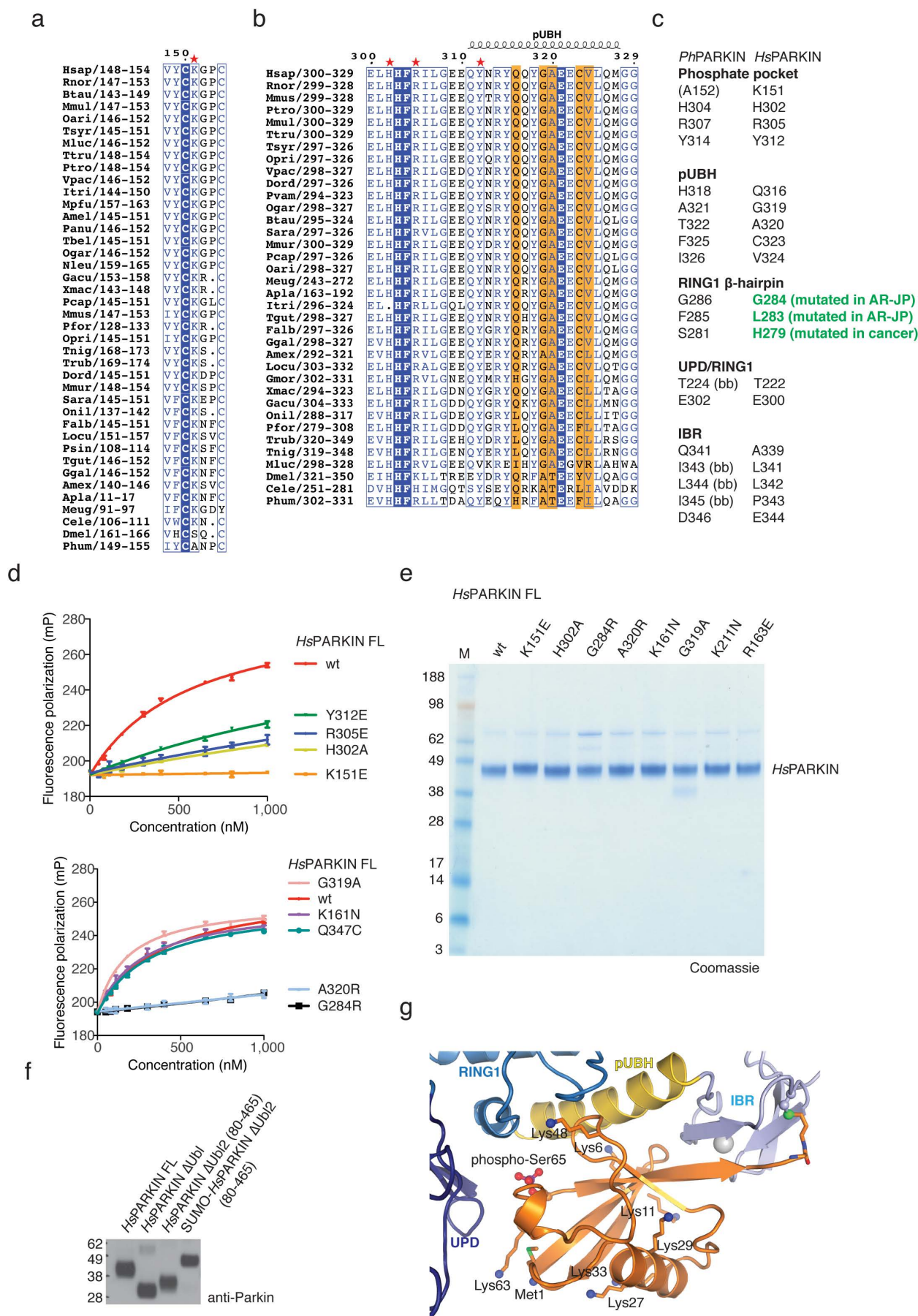


e



Extended Data Figure 2 | *Ph*PARKIN similarity with human and rat PARKIN and map quality. **a**, Structure-based sequence alignment of *Hs*PARKIN from human (top), *Rn*PARKIN (middle) and *Pediculus humanus corporis* (*Ph*PARKIN, bottom). The domains are indicated in boxes coloured according to structural figures. *Hs*PARKIN and *Ph*PARKIN are 45% identical within their crystallized constructs. Secondary structure elements are shown for *Hs*PARKIN Ubl domain (PDB 1IYF (ref. 36)) and for *Hs*PARKIN core domain (4BM9 (ref. 14)). Red asterisks denote the phospho-Ser65 ubiquitin binding pocket, and yellow spheres the residues contacting phosphoUb (see Fig. 2). A black asterisk denotes the catalytic Cys in RING2. **b**, Stereo representation of the asymmetric unit of *Ph*PARKIN–pUb crystals,

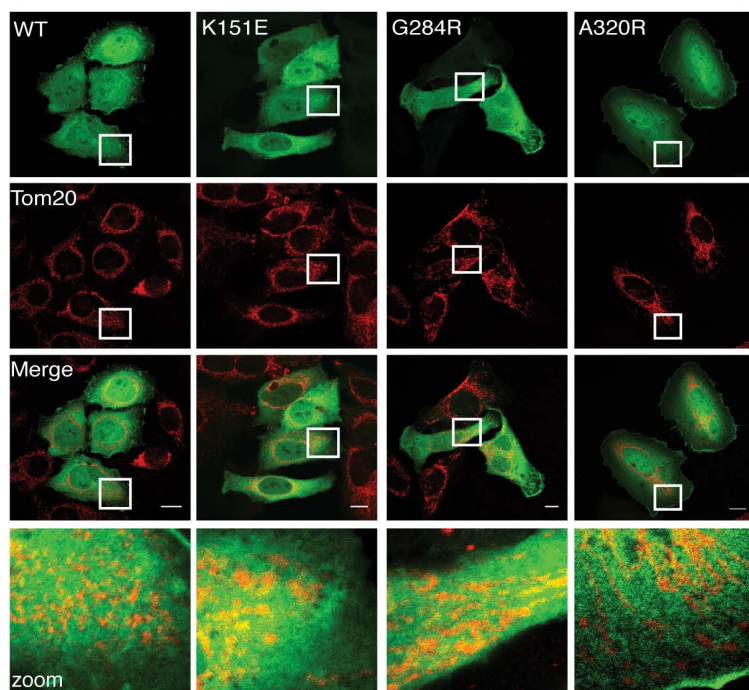
showing $2|F_o| - |F_c|$ electron density at 1σ , in blue for *Ph*PARKIN and in green for phosphoUb. **c**, Electron density detail, shown as in **b**, zooming in on phosphoUb phospho-Ser65, in stereo representation. **d**, Superposition of the two *Ph*PARKIN–pUb complexes in the asymmetric unit, coloured as in Fig. 1. The r.m.s.d. is 0.76 Å. Electron density is missing for parts of the flexible linker between IBR and RING2. **e**, Superposition of available PARKIN structures (PDB 4BM9 (ref. 14); 4I1H (ref. 15); 4K95 (ref. 16) and two *Ph*PARKIN–pUb complexes) in different colours, showing similar domain positions with respect to each other, with the exception of the IBR domain. Only the structure of full-length *Rn*PARKIN¹⁶ contains the Ubl domain.



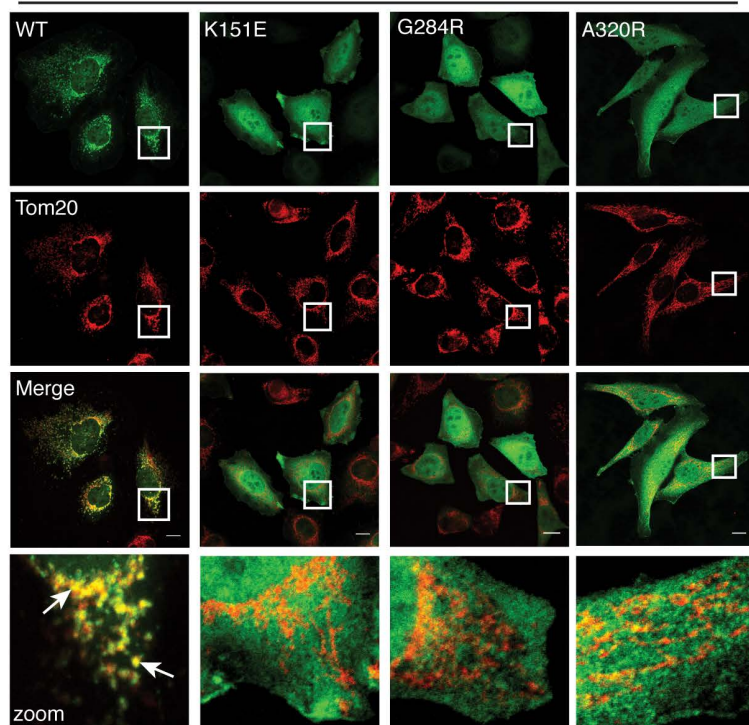
Extended Data Figure 3 | Conservation of phosphoUb interacting residues and biochemical analysis. **a, b**, An alignment of PARKIN from species available in Ensembl (<http://www.ensembl.org>) was curated by removing sequences with truncations or poorly sequenced regions. The residues involved in phosphoUb binding, comprising **(a)** the phosphate pocket and **(b)** the pUBH, are shown, and residues contacting phosphoUb are highlighted. **c**, Comparison of residues from *Ph*PARKIN and *Hs*PARKIN involved in phosphoUb binding. Highlighted in green is the *Hs*PARKIN β -hairpin residue Gly284 (Gly286 in *Ph*PARKIN), which is mutated to Arg in AR-JP. Other β -hairpin patient mutations, L283P (Phe285 in *Ph*PARKIN) and H279P (Ser281 in *Ph*PARKIN), introduce Pro residues, probably distorting the β -hairpin loop. The notation (bb) indicates backbone interaction. **d**, Fluorescence polarization assays performed with mutant PARKIN and phosphoUb. Assays from two independent experiments were combined to produce Fig. 2e. Measurements were performed in triplicate with error bars given as standard

deviation from the mean. **e**, Coomassie-stained SDS-PAGE gel showing *Hs*PARKIN proteins used in activity assays. **f**, Normalized proteins for PARKIN activity assays in Fig. 3d. **g**, Analogous to Fig. 2a, phosphoUb bound to *Ph*PARKIN is shown with Lys residues in stick representation, and free amine groups as blue spheres. Several Lys side chains (Lys29, Lys33, Lys48 and Lys63) were disordered in the electron density maps, suggesting high flexibility and solvent accessibility, and were modelled in their preferred side-chain rotamer for illustrative purposes. When bound to PARKIN, each Lys residue can be ubiquitinated, and the C terminus which is covalently attached to *Ph*PARKIN in the complex structure is probably more flexible and could also be attached to a more proximal (phospho)Ub in a chain. This indicates that PARKIN could interact with phosphoUb-containing polyubiquitin chains that were reported to be the PINK1 substrate on mitochondria^{7,8,11}. Molecular weight markers are in kDa in **e** and **f**.

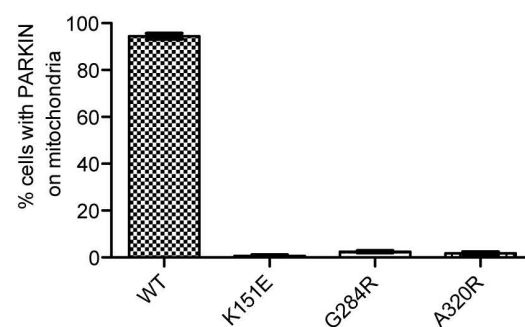
a

HeLa cells, YFP-*HsPARKIN*, DMSO

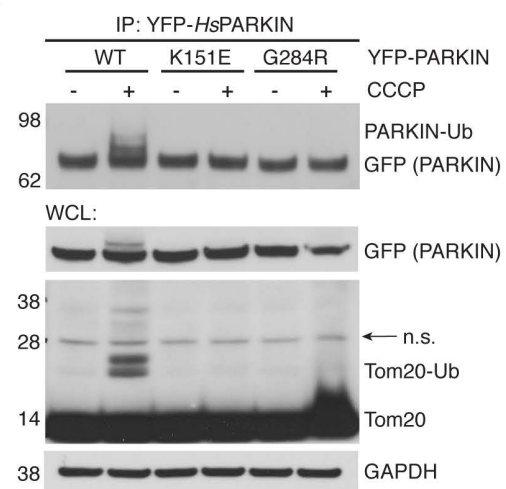
b

HeLa cells, YFP-*HsPARKIN*, 1h CCCP (10 μ M)

c



d

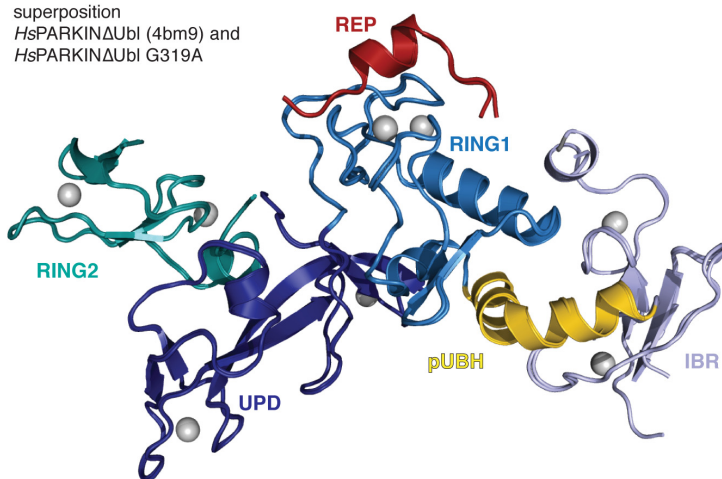


Extended Data Figure 4 | PARKIN phosphoUb-binding mutants do not translocate to mitochondria. **a, b,** HeLa cells that do not contain detectable levels of PARKIN were transiently transfected with YFP-*HsPARKIN* wild type and indicated mutants (green). Tom20 staining with anti-Tom20 antibody indicates mitochondria (red). Images are representative of three biological replicates. **a,** In DMSO-treated control cells, PARKIN does not co-localize with mitochondrial Tom20. **b,** After treatment with CCCP (10 μ M) for 1 h, wild-type PARKIN but not phosphoUb-binding mutants co-localize with Tom20. Scale bars, 10 μ m. Split channels are shown with overlay to illustrate co-localization. The YFP channel is identical to Fig. 2g. Digital zoom 10 \times . **c,** Quantification of cells with PARKIN localized at mitochondria in **b**, scored for 150 cells per condition in three biological replicates. PhosphoUb binding mutants do not show sustained mitochondrial localization. Error bars represent

standard error of the mean. **d,** HeLa cells transiently transfected with *HsPARKIN* wild type and mutants were treated with CCCP and YFP-*HsPARKIN* immunoprecipitated (see Methods). Whole-cell lysates (WCL) or immunoprecipitates (IP) were western blotted for PARKIN, Tom20 and GAPDH (loading control) as indicated. Ubiquitinated forms of PARKIN and Tom20 can be observed with wild-type PARKIN but not with PARKIN mutants. See also Fig. 2h where the A320R mutant was also included. These experiments were performed three times as biological replicates with similar results. Note that PARKIN autoubiquitination varied and was weaker in some experiments, while Tom20 ubiquitination was more robust. Molecular weight markers are in kDa. See Supplementary Information for uncropped blots.

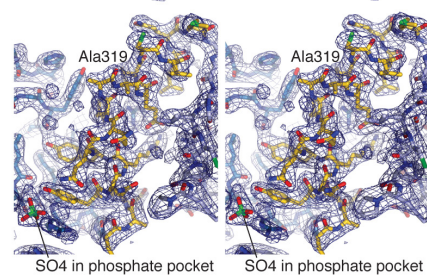
a

superposition
HsPARKINΔUbl (4bm9) and
HsPARKINΔUbl G319A



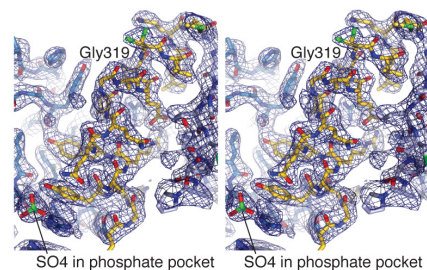
b

HsPARKINΔUbl G319A



c

HsPARKINΔUbl (4bm9)



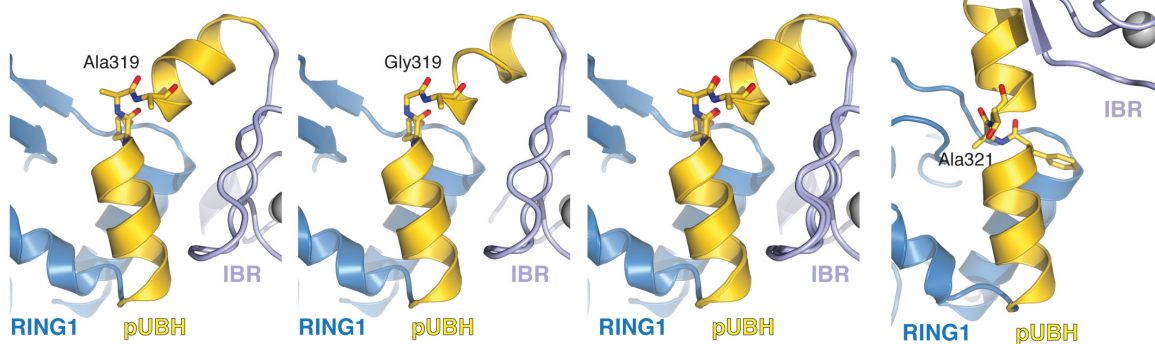
d

HsPARKINΔUbl G319A

wt HsPARKINΔUbl (4bm9)

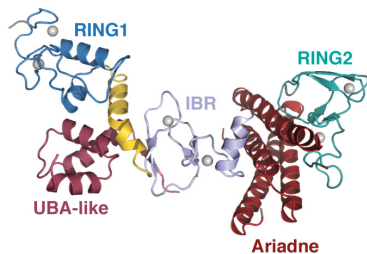
superposition

PhPARKIN-pUb (shown w/o pUb)



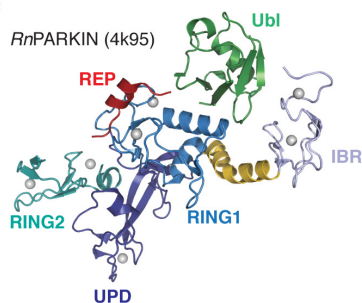
e

HHARI (4kbl)



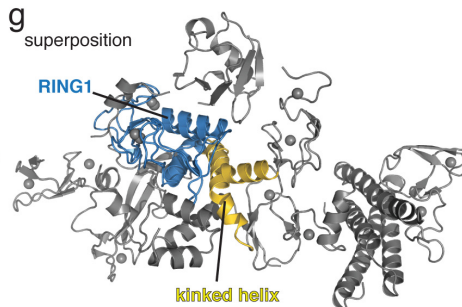
f

RnPARKIN (4k95)



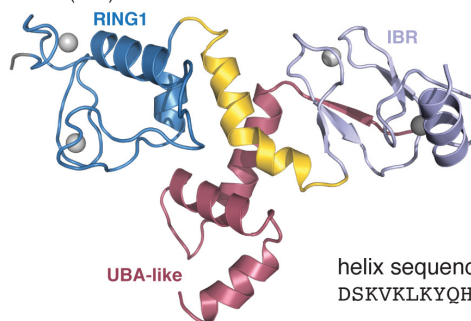
g

superposition



h

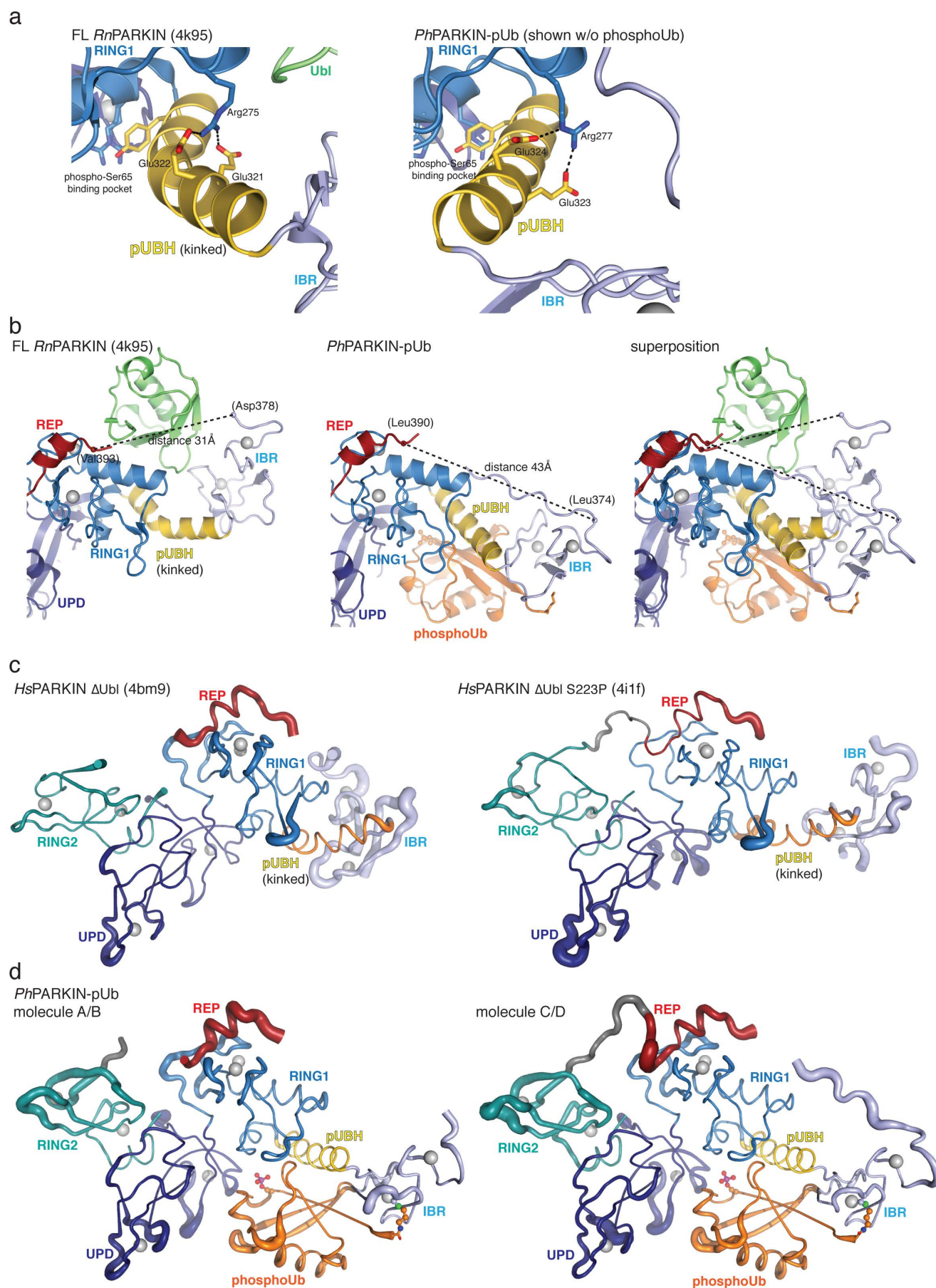
HHARI (4kbl)



helix sequence (kink):
DSKVCLKYQHLLTNSFVECNR

Extended Data Figure 5 | The kink in the pUBH. a–d, To understand whether the straight pUBH in *PhPARKIN* was a consequence of a helix-favouring mutation of Gly319 (*HsPARKIN*) to Ala (Ala321 in *PhPARKIN*), the *HsPARKIN*(Δ Ubl) G319A mutant was crystallized and a structure determined at 2.35 Å resolution (see Extended Data Fig. 1e and Methods). **a,** Superposition of *HsPARKIN*(Δ Ubl) G319A and wild-type *HsPARKIN*(Δ Ubl) (PDB 4BM9 (ref. 14)). The structures are virtually identical, both containing a kinked helix. **b, c,** Electron density detail of the kinked helix for the mutant (**b**) and wild type (**c**), shown in stereo representation, with $2|F_o| - |F_c|$ density contoured at 1σ . The C β atom of Ala319 is clearly defined in electron density and does not induce helix straightening in this crystallographic setting. **d,** Comparison of pUBH helices in *HsPARKIN*(Δ Ubl) G319A (left), *HsPARKIN*(Δ Ubl) wild type (second from left), superposition of the two (second from right), and for comparison the straight pUBH in *PhPARKIN*-pUb. **e, f,** Structure of the RBR

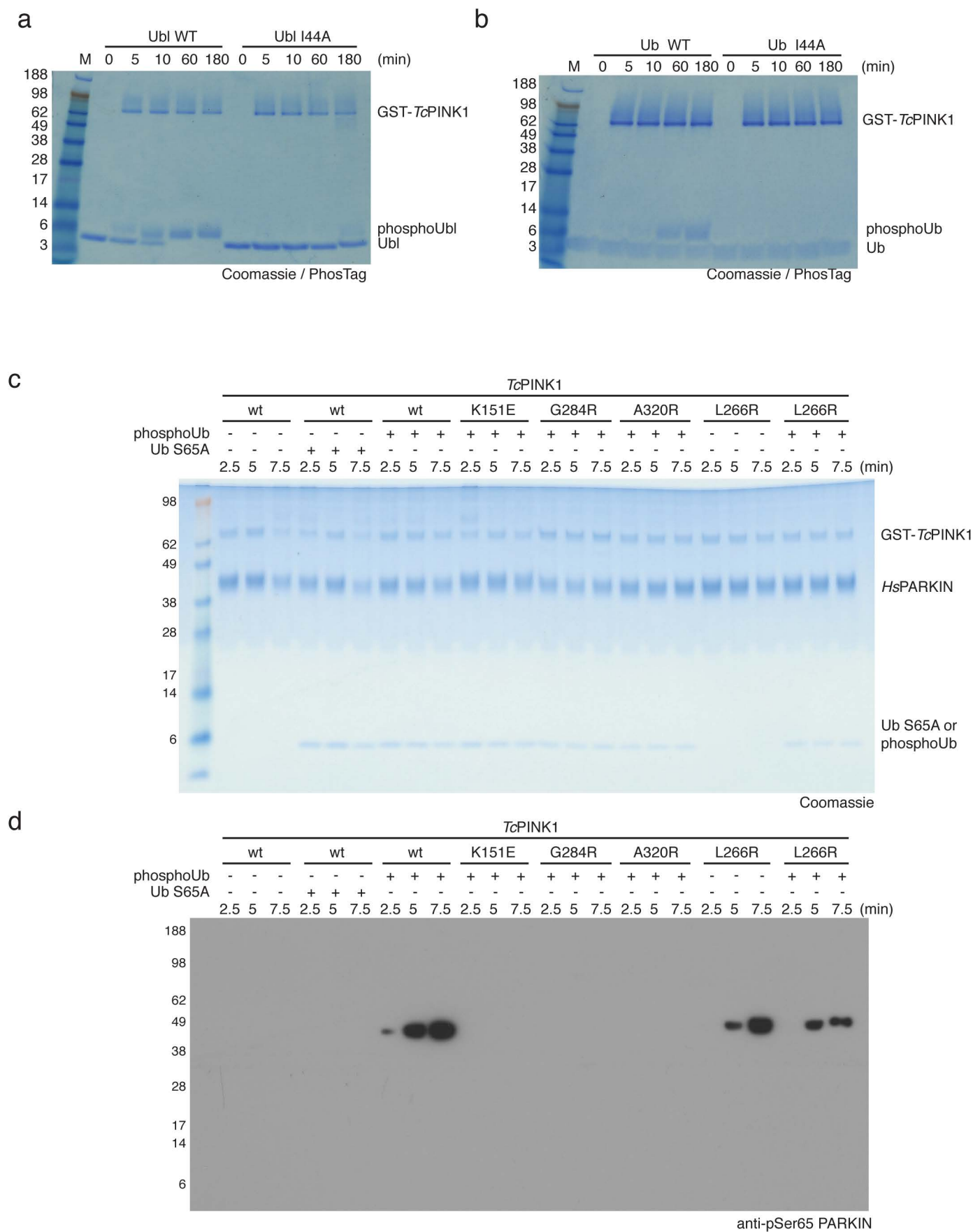
E3 ligase HHARI (**e**) in an autoinhibited form (PDB 4KBL (ref. 23)) showing an entirely different RBR module as compared to full-length *RhPARKIN* (PDB 4K95 (ref. 16)) in **f**. RING1, IBR, RING2 domains are coloured as for PARKIN in Fig. 1, and other domains (UBA-like domain and Ariadne domain in HHARI) are coloured in red. Interestingly, the pUBH equivalent helix in HHARI is kinked at a similar position as compared to mammalian PARKIN. **g,** Superposition of structures from **e** and **f** on their RING1 domains. In HHARI, the pUBH-equivalent helix kinks in a different manner as compared to *HsPARKIN*. **h,** In HHARI, the kinked helix seems to be stabilized by an interaction in *cis* with the UBA-like domain that binds NEDD8 (ref. 37). The sequence of the helix does not contain Gly residues, but is kinked at a Thr residue (Thr263). It will be interesting to see whether helix straightening occurs in active forms of HHARI that can be induced by binding NEDD8-modified cullins³⁷.



Extended Data Figure 6 | Structural detail and B-factor analysis.

a, PhosphoUb-induced pUBH straightening is energetically neutral as the two hydrogen bonds between helix residues and the RING1 core are not lost but only adjusted. Left: full-length *Rn*PARKIN structure (PDB 4K95 (ref. 16)). Right: *Ph*PARKIN–pUb structure. **b**, Structure of full-length *Rn*PARKIN (PDB 4K95 (ref. 16), left), *Ph*PARKIN–pUb (middle), and a superposition of the two (right), in which C α atoms of structurally identical, ordered residues of the IBR–REP linker are shown as spheres. The distance between these residues is indicated by a dotted line. The linker sequence is highlighted in Extended Data Fig. 2. **c**, Structures of PARKIN in which B-factors were refined for individual atoms (PDB 4BM9 (ref. 14); 4I1H (ref. 15)) are shown in a ribbon

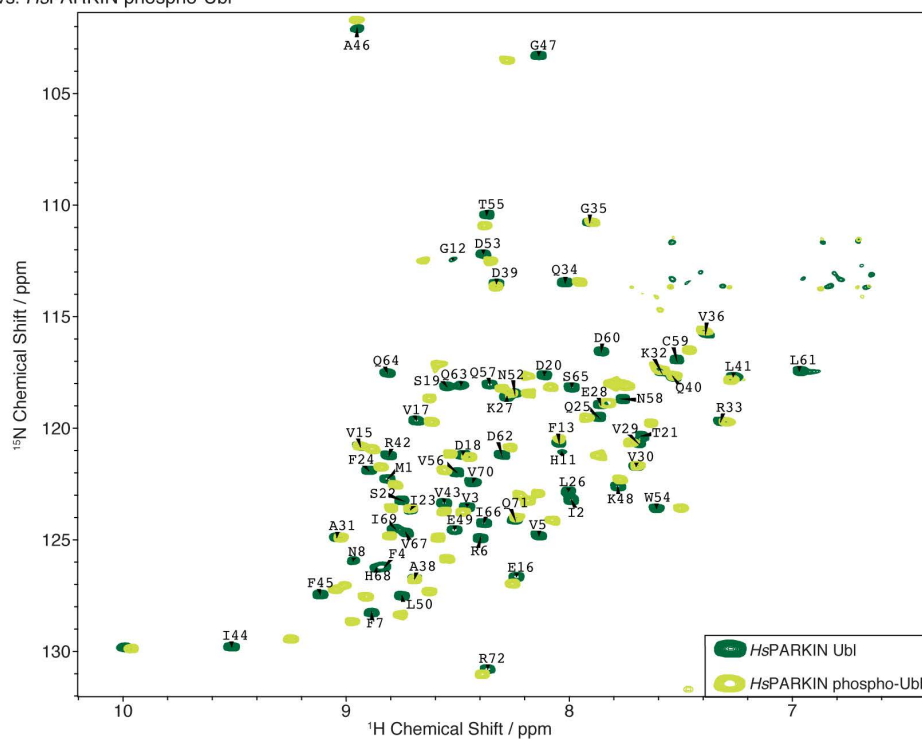
representation, in which the ribbon thickness indicates B-factors differences, with thin ribbons indicating low and thick ribbon indicating high B-factors. The full-length *Rn*PARKIN structure (PDB 4K95 (ref. 16)) is not included as in this 6.5 Å structure overall B-factors were assigned. **d**, Structures of *Ph*PARKIN–pUb are shown as in **c** for each molecule of the asymmetric unit. The REP and RING2 elements are destabilized as indicated by higher B-factors. The rigid core of the protein has shifted from the UPD–RING1–RING2 interfaces (compare **c**) to the UPD–RING1–IBR–phosphoUb interfaces. B-factor analysis may be distorted by neighbouring molecules and crystal contacts, which are not indicated here.



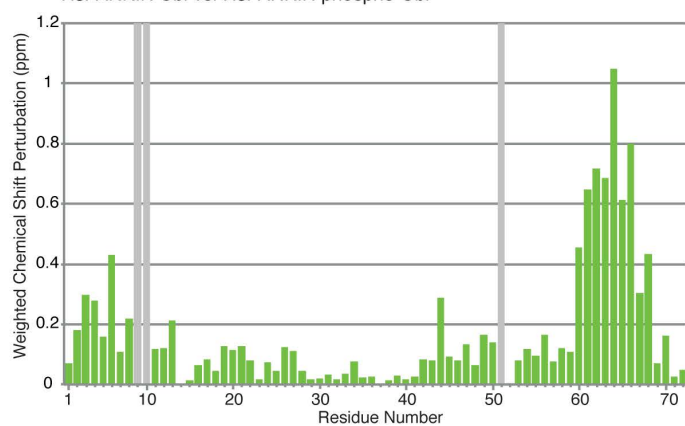
Extended Data Figure 7 | The Ile44 patch is essential for PINK1-mediated phosphorylation of Ub and PARKIN Ubl. All assays were performed three times with consistent results. **a, b**, Coomassie-stained PhosTag gels comparing the phosphorylation of **(a)** the *HsPARKIN* Ubl domain (amino acids 1–72) and **(b)** ubiquitin. In both cases, the wild-type form is compared with the I44A mutant form of the protein. GST–*TcPINK1* does not efficiently phosphorylate

the I44A mutants of ubiquitin or of the *HsPARKIN* Ubl domain. This is important since the Ile44 patch in the PARKIN Ubl domain is inaccessible and binds to RING1 in the structure of full-length *RnPARKIN* (PDB 4K95 (ref. 16)) (see Fig. 4a). **c, d**, Controls for Fig. 4c. Coomassie-stained gel with proteins labelled **(c)**, and full-size blot **(d)** for Fig. 4c. Molecular weight markers are in kDa for all panels.

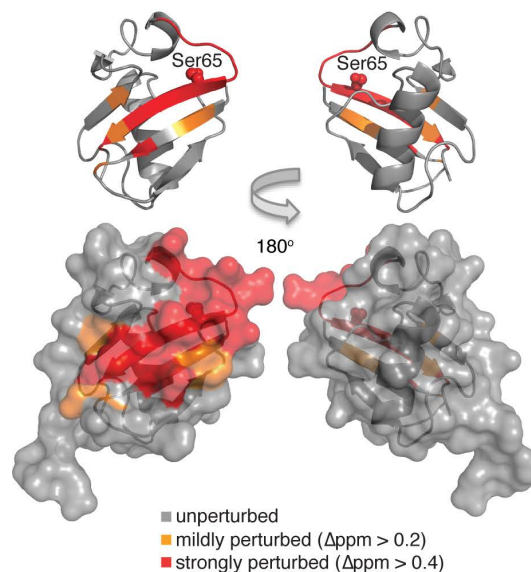
a *HsPARKIN* Ubl vs. *HsPARKIN* phospho-Ubl



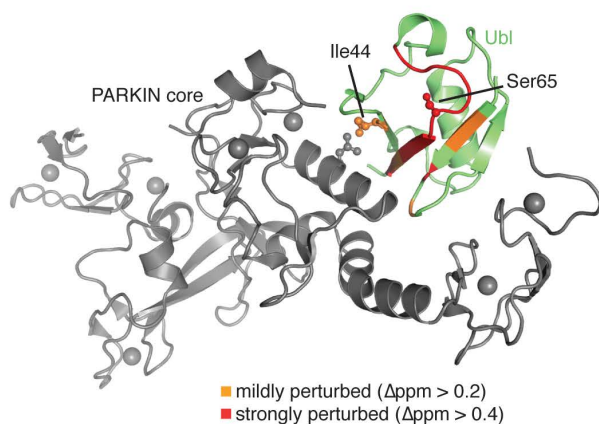
b *HsPARKIN* Ubl vs. *HsPARKIN* phospho-Ubl



c *HsPARKIN* Ubl (1iyf)



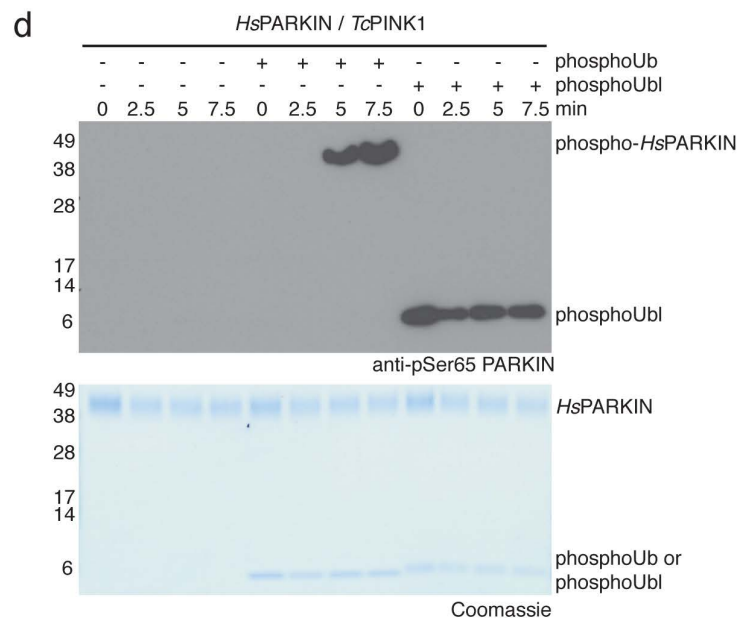
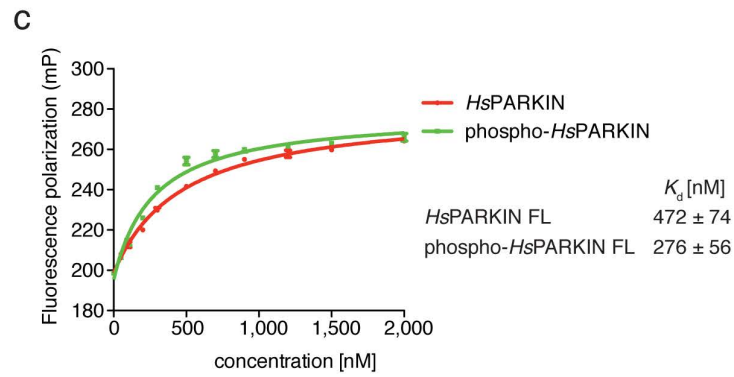
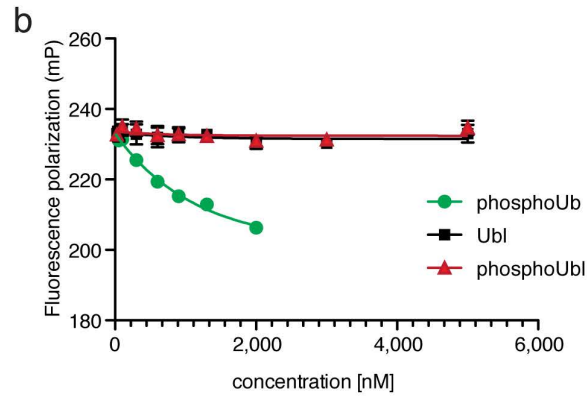
d



Extended Data Figure 8 | NMR analysis of phosphorylated Ubl. **a**, BEST-TROSY spectra for isotope-labelled *Hs*PARKIN Ubl domain (dark green) and phospho-Ubl domain (light green) with resonances assigned for the Ubl domain. **b**, Chemical shift perturbation of Ubl with respect to phospho-Ubl, showing significant perturbations in the region of phosphorylation (Ser65), the last β -strand and neighbouring β -strands. Grey bars, exchange broadened resonances. **c**, Mapping of perturbed resonances onto the previously

determined NMR structure of *Hs*PARKIN Ubl (PDB 1IYF (ref. 36)). The perturbed residues cluster in the Ser65-containing loop and in proximity to the Ile44 patch of the Ubl. **d**, Mapping of the perturbed resonances to the structure of *Rn*PARKIN (PDB 4K95 (ref. 16)) shows that they perturb the interface between the Ubl domain and the PARKIN core. Thus, phosphorylated Ubl may not be able to (re)bind PARKIN at the same binding site.

a Scenario 1 - phosphoUbl competes with phosphoUb

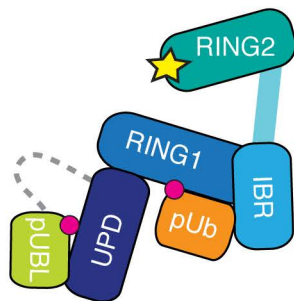


Extended Data Figure 9 | PARKIN phosphoUbl does not compete with phosphoUb. **a**, A possible scenario is that the phosphoUbl competes with phosphoUb for the phosphoUb binding site on PARKIN. This could be favoured since the interaction would occur in *cis*. **b**, Fluorescence polarization competition experiment increasing the concentration of Ubl, phosphoUbl or phosphoUb with respect to full-length HsPARKIN in the presence of FLAsH-tagged phosphoUb. Measurements were performed in triplicate with error bars given as standard deviation from the mean. While unlabelled phosphoUb competes with labelled phosphoUb in the reaction, Ubl or phosphoUbl do not compete with phosphoUb. **c**, Binding of FLAsH-phosphoUb to HsPARKIN and phospho-HsPARKIN. The measurements were performed in the same

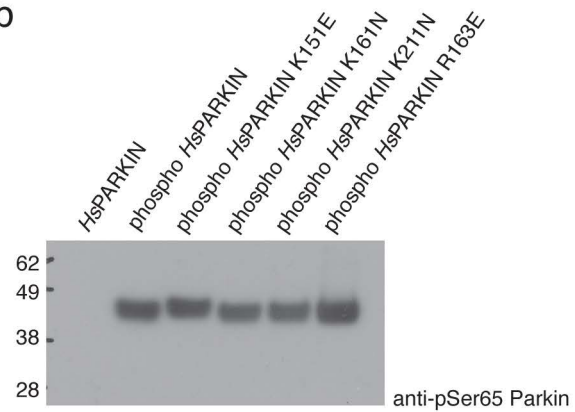
experiment as samples in Fig. 2d. If phosphoUbl interacts with HsPARKIN in *cis*, phosphoUb binding should be inhibited. In contrast, binding of phosphoUb to phospho-PARKIN is slightly enhanced as also reported in ref. 7. Measurements were performed in triplicate and error bars represent standard deviation from the mean. **d**, PARKIN phosphorylation assays as in Fig. 4c, including phosphoUbl as well as phosphoUb (both at 10 μ M). While addition of phosphoUb induces PARKIN phosphorylation, addition of phosphoUbl does not, indicating that the Ubl domain is not released from the PARKIN core. The experiment has been performed three times with consistent results. Molecular weight markers are in kDa.

a

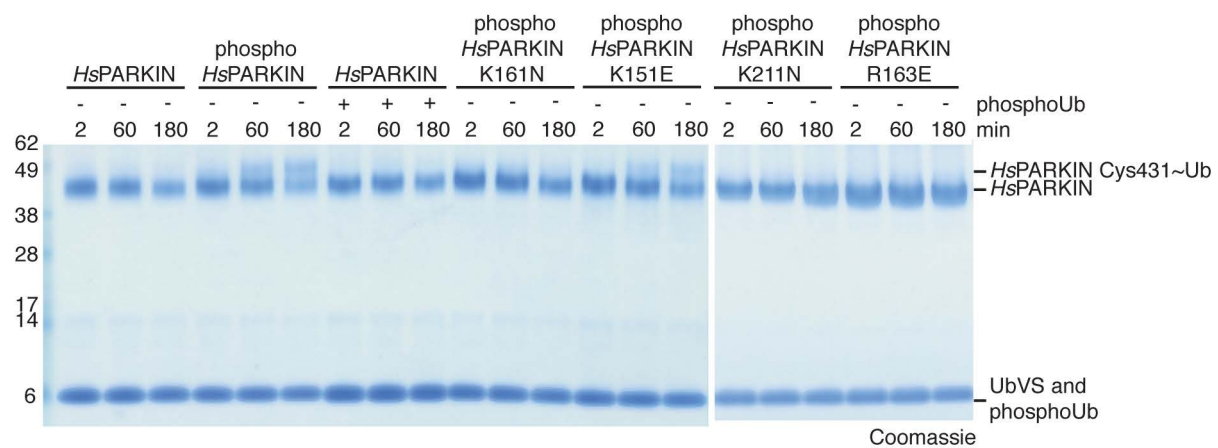
Scenario 2 - phosphoUbl binds elsewhere on PARKIN



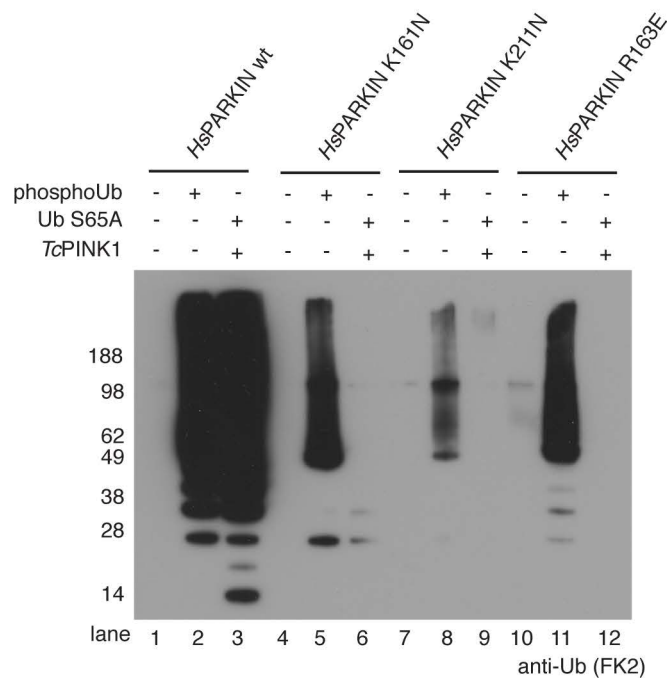
b



c



d



Extended Data Figure 10 | PhosphoUbl-induced ‘opening’ of PARKIN relies on a phosphate pocket in the UPD. **a**, A second scenario would be that the phosphorylated Ubl domain rebinds to PARKIN at an alternative site. We previously speculated that the UPD contains a phosphate-binding site that is lined by two AR-JP patient mutations, K161N and K211N, and also contains Arg163 (ref. 14). **b**, Full-length *HsPARKIN* mutants as indicated were phosphorylated *in vitro* with GST-*PhPINK1*, resolved on SDS-PAGE and western blotted using an anti-pSer65 PARKIN antibody (Abcam cat no. ab154995). These proteins were used in **c**. **c**, Ub-vinyl sulfone (Ub-VS) modification of the active site Cys431 of *HsPARKIN* and *HsPARKIN* mutants with and without phosphoUb from a time course experiment is assessed on Coomassie-stained gels. Ub-VS reacts with ‘open’ forms of PARKIN, and was previously shown to modify phospho-PARKIN but not phosphoUb-activated PARKIN⁷. Phospho-PARKIN-mediated opening depends on the phosphate pocket present in the UPD, since phospho-PARKIN K211N, K161N (two AR-JP patient mutations) and R163E abrogated or impaired

modification by Ub-VS and do not appear to have an accessible catalytic Cys, while the phosphorylated phosphoUb-binding-deficient mutant K151E is readily modified. The experiment was performed two times with consistent results and gels have been collated from two different assays (indicated by the gap). **d**, PARKIN ubiquitination reactions in presence of E1, UBE2L3, ubiquitin or ubiquitin S65A, ATP and GST-*TcPINK1* for 2 h. PARKIN is activated by phosphoUb or by PARKIN Ubl phosphorylation in absence of phosphoUb (with ubiquitin S65A) (lanes 2/3). Mutants in the UPD phosphate pocket can still be activated by phosphoUb (albeit not to the same extent) (lanes 5, 8, 11) but are inactive when the Ubl domain is phosphorylated (lanes 6, 9, 12). This could suggest that the phosphoUbl binds back to the PARKIN UPD pocket. However, we cannot exclude the possibility that, for example, the linker between the Ubl and the UPD plays a more active role in PARKIN activation. The experiment was performed three times with consistent results. Molecular weight markers are in kDa for **b–d**.

Propagation of conformational changes during μ -opioid receptor activation

Rémy Sounier¹, Camille Mas¹, Jan Steyaert^{2,3}, Toon Laeremans^{2,3}, Aashish Manglik⁴, Weijiao Huang⁴, Brian K. Kobilka⁴, Hélène Déméné⁵ & Sébastien Granier¹

μ -Opioid receptors (μ ORs) are G-protein-coupled receptors that are activated by a structurally diverse spectrum of natural and synthetic agonists including endogenous endorphin peptides, morphine and methadone. The recent structures of the μ OR in inactive¹ and agonist-induced active states (Huang *et al.*, ref. 2) provide snapshots of the receptor at the beginning and end of a signalling event, but little is known about the dynamic sequence of events that span these two states. Here we use solution-state NMR to examine the process of μ OR activation using a purified receptor (mouse sequence) preparation in an amphiphile membrane-like environment. We obtain spectra of the μ OR in the absence of ligand, and in the presence of the high-affinity agonist BU72 alone, or with BU72 and a G protein mimetic nanobody. Our results show that conformational changes in transmembrane segments 5 and 6 (TM5 and TM6), which are required for the full engagement of a G protein, are almost completely dependent on the presence of both the agonist and the G protein mimetic nanobody, revealing a weak allosteric coupling between the agonist-binding pocket and the G-protein-coupling interface (TM5 and TM6), similar to that observed for the β_2 -adrenergic receptor³. Unexpectedly, in the presence of agonist alone, we find larger spectral changes involving intracellular loop 1 and helix 8 compared to changes in TM5 and TM6. These results suggest that one or both of these domains may play a role in the initial interaction with the G protein, and that TM5 and TM6 are only engaged later in the process of complex formation. The initial interactions between the G protein and intracellular loop 1 and/or helix 8 may be involved in G-protein coupling specificity, as has been suggested for other family A G-protein-coupled receptors.

μ ORs are membrane proteins belonging to the G-protein-coupled receptor (GPCR) superfamily that are responsible for the analgesic and addictive properties of morphine⁴, a natural alkaloid contained in opium⁵. In fact, due to the clinical importance of opioids in pain management and addiction, five decades of medicinal chemistry generated a very chemically diverse opioidergic ligand system (reviewed in ref. 6) that can modulate the μ OR intracellular signalling, namely the inhibitory heterotrimeric G_i protein and arrestin pathways⁷. The pharmacology of these different compounds are very well characterized and they present a wide spectrum of activity with respect to efficacy towards G protein and/or arrestin from very weak partial agonism to full agonism^{7,8}. Evidence suggests that activation of the arrestin pathway may be responsible for some of the adverse clinical effects of the opioids drugs currently on the market⁷.

Intrinsic efficacy is a parameter used to define the effect of a ligand on GPCR activation of specific signalling pathways⁹. To understand intrinsic efficacy, it is necessary to understand the process of GPCR activation. So far this process has only been described for three GPCRs using biophysical and crystallographic analyses of both inactive and active states: rhodopsin¹⁰, the β_2 -adrenergic receptor (β_2 AR)^{11,12} and

muscarinic M2 receptor (M2R)¹³. The most complete characterization of the GPCR activation process by diffusible ligands came from NMR studies on the β_2 AR^{14–18}, showing that the binding of agonist alone is associated with conformational heterogeneity in the receptor and is not sufficient to stabilize the active β_2 AR conformation. Whether this observation is generalizable to other GPCRs has not been determined.

Crystal structures are now available for the μ OR in both inactive and fully agonist-induced states, representing the beginning and end of a signalling event (Huang *et al.*, ref. 2). Here, in an effort to better understand the process of μ OR activation upon binding of opioid agonists, we took advantage of the versatility of NMR spectroscopy (as recently described for the β_2 AR¹⁹) to analyse the activation signal propagation occurring in different receptor domains (see Extended Data Fig. 1 and Methods for further technical details).

Dimethylamines give peaks with different intensities in the ¹H-¹³C heteronuclear multiple quantum coherence (HMQC) spectrum of unliganded ¹³C-dimethylated μ OR (¹³Cm2- μ OR) preparations (Fig. 1a) that we unambiguously assigned to ten lysines and to the shortened N terminus (starting at G52) using systematic mutagenesis and proteolysis experiments (Extended Data Fig. 2). It is interesting to note the coexistence of broad and irregular peaks (that is, K98^{ICL1}, K100^{ICL1} and K344^{ICL1} (ICL1, intracellular loop 1)) with sharp and intense signals (K209^{ECL2}, K269^{ECL2} (ECL2, extracellular loop 2)) in the apo-state. Because a peak line width directly reports on the homogeneity of the local fluctuating magnetic field which depends on the dynamics of the labelled domain, this observation qualitatively indicates that the different domains under investigation (that is, TM6 and ICL1–helix 8 (H8)) present different dynamic properties (in the μ s to ms range). Indeed, broad peaks suggest that the ε -N[(¹³CH₃)₂] moieties detect two or more conformations (environments) that exchange on an intermediate time scale (low ms), while sharp peaks may represent a single conformation or two or more conformations that exchange on either a faster or slower time scale. The absence of a signal for a specific lysine may represent a lack of side chain dynamics or a dynamic time-scale in the ‘slow intermediate’ regime²⁰. We have focused our analysis on a specific set of NMR correlation peaks that were separated in two groups of sensors: the extracellular (G52^{N-ter} and K209^{ECL2}) and intracellular (K98^{ICL1}, K100^{ICL1}, K260^{ICL1}, K269^{ECL2}, K271^{ECL2} and K344^{ICL1}) lysines (Fig. 1b–d). The signals for other probes are very weak (K174^{ICL2}, K233^{ICL2}, K303^{ICL2}) or even not observable (K141^{ICL2} and K185^{ICL2}), and spectral modifications are difficult to interpret (Extended Data Fig. 2).

We first analysed the effect of BU72, a high-efficacy μ OR agonist²¹, on the NMR signals using a saturating amount of ligand and/or a saturating amount of a G protein mimetic nanobody (Nb33) (Fig. 2). We selected BU72 because of its extremely high efficacy/potency and its slower dissociation rate relative to other agonists. Moreover, this ligand was successfully used in the crystallization studies (Huang *et al.*, ref. 2). Nb33 is closely related to Nb39, the nanobody

¹Institut de Genomique Fonctionnelle, CNRS UMR-5203 INSERM U1191, University of Montpellier, F-34000 Montpellier, France. ²Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. ³Structural Biology Research Center, VIB, Pleinlaan 2, B-1050 Brussels, Belgium. ⁴Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁵Centre de Biochimie Structurale, CNRS UMR 5048-INSERM 1054- University of Montpellier, 29 rue de Navacelles, 34090 Montpellier Cedex, France.

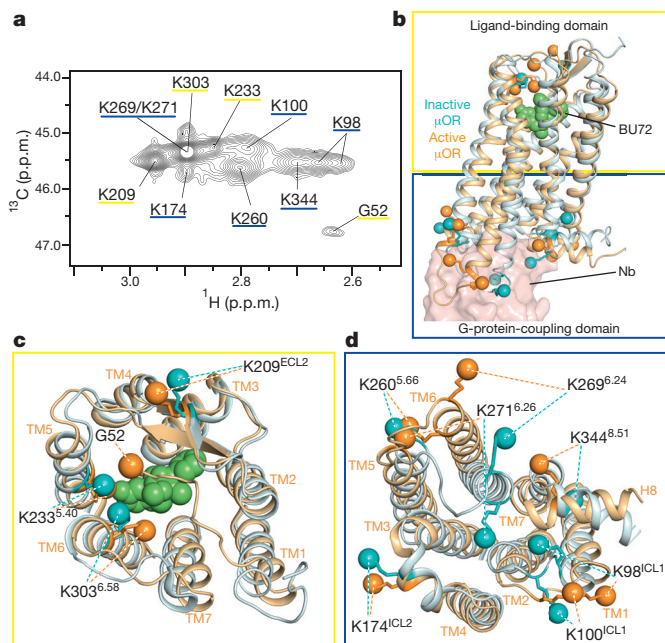


Figure 1 | HMQC spectrum of unliganded ^{13}C -dimethylated μOR and positions of lysine residues in the inactive (cyan blue) and active (orange) μOR . **a**, Position and identity of eleven correlated ^1H - ^{13}C peaks corresponding to the μOR dimethyllysines and amino terminus. **b**, Superposition of inactive (cyan blue, 4DKL) and active (orange, 5C1M) conformation of μOR highlighting the lysines (sticks) and the $\varepsilon\text{-NH}_2$ (spheres) positioned in the extracellular face (yellow box) and in the intracellular face (blue box) of the μOR . Also shown are the BU72 ligand (green spheres) and the G protein mimetic nanobody position (salmon surface). **c**, **d**, Extracellular (**c**) and intracellular (**d**) view of the μOR showing the reorganization of the lysine side chains upon receptor activation from inactive (cyan spheres) to fully active (orange spheres). The shortened amino-terminal domain (N terminus) starting at G52 was not modelled in the inactive state structure but makes contact with the 7TM bundle to cap the binding pocket in the agonist-induced state.

used to stabilize and capture the μOR agonist-induced state in crystallization studies (Huang *et al.*, ref. 2). Nb33 and Nb39 differ by only two amino acids in a region that does not directly interact with the receptor, and they have similar G-protein-like effects on the affinity of agonist for the μOR . They also have similar effects in NMR experiments (Extended Data Fig. 3). We used Nb33 throughout this study because it can be purified in much larger quantities than Nb39 (see Methods).

In the extracellular surface, the peak corresponding to K209^{ECL2} was slightly modified upon treatment with BU72 (Fig. 2a, LR panel) and, as better observed in the ^1H dimension, was decreased in intensity and shifted downfield (Fig. 2b, ECL2 panel). As discussed above, this spectral modification indicates a change in the conformational state of the μOR , probably as a result of a modification of the K209^{ECL2} micro-environment, as the β -hairpin containing this residue undergoes a change upon receptor activation (Fig. 1c). When the sample was treated with a saturating concentration of the Nb33 only, no significant changes were observed (Fig. 2a, RNb panel), suggesting an absence of environment changes around K209^{ECL2} in the absence of agonist. When both BU72 and Nb33 were added to the $^{13}\text{C}_2\text{-}\mu\text{OR}$, the K209^{ECL2} peak intensity decreased further as compared to the sample treated with BU72 alone (Fig. 2a, LRNb panel and Fig. 2b, ECL2 panel). We also observed changes in the N terminus upon binding of BU72 alone; however, in this case the peak intensity increases. BU72 and Nb33 co-treatment slightly increased the effect observed for BU72 alone (Fig. 2b, N-terminal panel). It is of interest that the active-state crystal structure of the μOR reveals that the N terminus folds into the binding pocket and interacts with BU72 (Huang *et al.*, ref. 2).

The potentiating effects of Nb33 on BU72 are much more pronounced in the intracellular part of $^{13}\text{C}_2\text{-}\mu\text{OR}$, as BU72 or Nb33

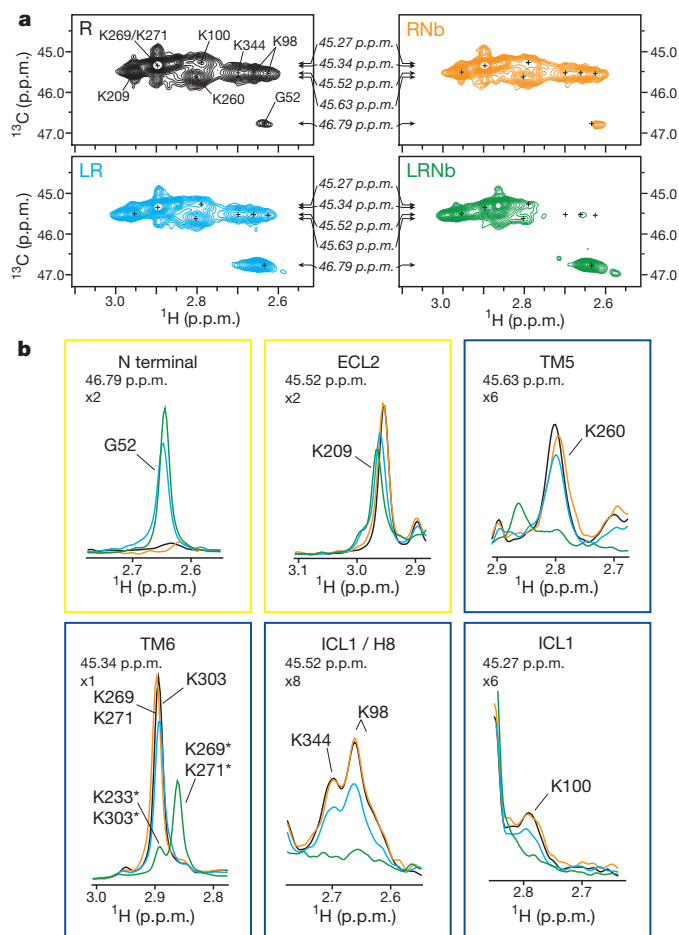


Figure 2 | Activation of μOR by BU72. **a**, HMQC spectra of unliganded ^{13}C -dimethylated μOR (30 μM , R, black), ^{13}C -dimethylated μOR treated with a saturating concentration of BU72 (150 μM , LR, blue), ^{13}C -dimethylated μOR treated with BU72 and a saturating concentration of Nb33 (60 μM , LRNb, green) and of unliganded ^{13}C -dimethylated μOR treated with a saturating concentration of Nb33 (60 μM , RNb, orange). The small crosses indicate the position of peaks visible in the apo-state that are reduced in intensity or no longer visible in the other spectra. Arrows indicate the position of the 1D slice represented in **b**. **b**, 1D slice of HMQC spectra in the ^1H dimension highlighting the effect of BU72 alone (blue) and BU72 + Nb33 (green) on the N terminus (G52) and on the indicated set of dimethyllysines (ECL2, TM5, TM6, ICL1 and H8 lysines). The yellow and blue boxes indicate the positions in the extracellular and intracellular domains, respectively. The zoom level for each peak and the ^{13}C chemical shifts are indicated in the top left of each panel. Asterisks indicate the chemical shifts of methyl probes in the presence of both BU72 and Nb33.

induced only small spectral changes by themselves (Fig. 2a, LR and RNb panels), while formation of the ternary complex led to more dramatic spectral modifications (Fig. 2a, LRNb panel). The proton peak centred around 2.9 p.p.m., which corresponds to the lysines K269^{6.24}, K271^{6.26} and K303^{6.58} positioned in TM6, almost disappeared (the leftover peak being K303^{6.58}, see assignments in Extended Data Fig. 2) with a concomitant appearance of an intense peak at 2.85 p.p.m. due to K269^{6.24} and K271^{6.26} alone (Fig. 2b, TM6 panel). This spectral shift, which is not observed in spectra from receptor treated with BU72 alone, is consistent with the 10 Å outward movement of TM6 upon receptor activation that repositions the K269^{6.24} and K271^{6.26} side chains (Fig. 1d). The K260^{5.66} peak intensity is also markedly decreased by exposure to both BU72 and Nb33, and thus represents a sensor for the receptor conformational change in TM5 (Fig. 2b, TM5 panel). These results show that binding of the high-affinity agonist BU72 alone is not sufficient to fully stabilize the

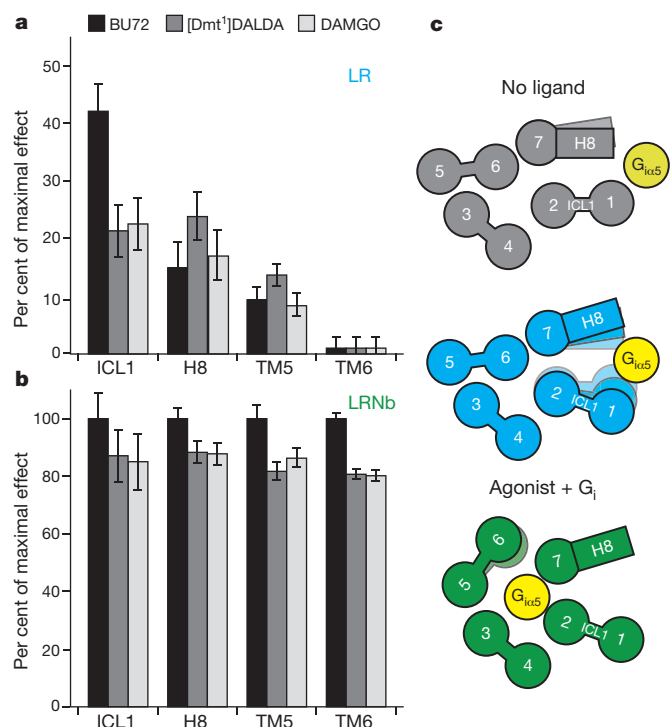


Figure 3 | Quantitative analysis of spectral changes in intracellular domains of μ OR and a structural interpretation. **a, b**, Percentage of maximal effect in the volume of indicated peaks in sample treated with agonist alone (BU72, [Dmt¹]DALDA and DAMGO) (**a**) or agonists together with Nb33 (**b**), highlighting the stronger effect of agonist in absence of Nb33 on the ICL1 and H8 domains as compared to TM5 and TM6 domains. Error bars represent the uncertainty in volume determination due to random noise (see Methods). **c**, Model proposing the propagation of conformational changes from the agonist-binding pocket to the ICL1 and H8 domains initiating interactions with the G protein G_i (represented by the α 5 C-terminal domain of G_i, yellow circle) (middle panel) before the fully agonist-induced state of the μ OR is reached (bottom panel), with the direct interaction of the G_i with TM5 and TM6. Numbered circles represent the identity and the position of the TM domains viewed from the intracellular side of the membrane.

active conformations of TM5 and TM6 and support the notion of a weak connection between the ligand-binding and the G-protein-coupling domains of the μ OR. This weak coupling was previously observed for the β_2 AR, and was thoroughly investigated more recently using NMR and DEER spectroscopy³. Of interest, the results obtained with the β_2 AR and the μ OR stand in contrast to rhodopsin, for which a stronger link between photoisomerization of retinal and structural changes at the cytoplasmic end of TM6 has been observed²².

In addition to the TM5 and TM6 probes, we also characterized the conformational transitions in the ICL1 (K98^{ICL1} and K100^{ICL1}) and H8 domains (K344^{H8}) of μ OR that also undergo conformational changes upon μ OR activation (Fig. 1d). Comparison of the R panel with the LRNb panel in Fig. 2a revealed a near complete loss of intensity of K98^{ICL1}, K100^{ICL1} and K344^{H8} peaks upon μ OR treatment with BU72 and Nb33 (Fig. 2a, b, ICL1/H8 panels, green trace). While a complete loss of intensity for these peaks was only observed with both BU72 and Nb33, we did observe a reduction in intensity with BU72 alone (Fig. 2b, ICL1/H8 panels, blue trace). The loss of peak intensity could originate from a change in the local dynamic of ICL1 and H8 domains in the μ s to ms timescale that broadens the correlation peaks beyond the detection limit, a well characterized phenomenon in NMR spectroscopy²³, and suggests that receptor activation is markedly modifying the structural dynamic of these domains. It should be noted that none of the lysines discussed here directly interact with the related Nb39 in the crystal structure. Changes in the NMR signal intensity

therefore reflect the μ OR activation process. Moreover, the effects of agonist and Nb33 can be reversed by treatment with the antagonist naloxone (Extended Data Fig. 4).

Taken together our data demonstrate that binding of the G protein surrogate is required for the agonist to fully stabilize the active conformation in both extracellular and intracellular domains, suggesting the existence of a two-way allosteric coupling between the extracellular μ OR ligand-binding domain and the G-protein-coupling interface. These data are in agreement with what has been observed with the isotopically labelled β_2 AR¹⁴, and suggest that this behaviour could be a common feature of the GPCR activation process.

We next quantitatively analysed the effect of BU72 alone (LR) and together with Nb33 (LRNb) on the volumes of the different intracellular lysine peaks (from TM5, TM6, ICL1 and H8 domains). Of particular interest, we found a notable difference in the effect of BU72 alone on TM5 and TM6 compared to its effect on ICL1 and H8 domains (Fig. 3a). While BU72 alone had a slight effect on TM5 and TM6 signal intensity, it produced larger spectral changes for the ICL1 and H8 sensors (Fig. 3a), suggesting that the allosteric coupling from the agonist-binding pocket to ICL1/H8 is stronger than the coupling to TM5 and TM6. This effect is not specific to BU72 as we observed the same phenomenon with two other agonists: [Dmt¹]DALDA (2',6'-dimethyltyrosine-D-Arg-Phe-Lys-NH₂) and DAMGO (Tyr-D-Ala-Gly-NMe-Phe-Gly-ol) (Fig. 3a and Extended Data Figs 5 and 6). Notably, the formation of the ternary complex leads to larger volume changes for the different domains under investigation (Fig. 3b) and, together with Fig. 3a, demonstrates that the fully agonist-induced state can only be achieved in the presence of an agonist and a G protein mimetic nanobody.

While we observe that agonists alone have larger effects on ICL1 and H8 reporters than on TM5 and TM6 reporters, our steady-state experiments do not provide information about the temporal sequence of these conformational changes; however, we speculate that a G protein may first engage ICL1 and/or H8 before engaging TM5 and TM6. In fact, the H8 domain is critical for G-protein coupling, as observed in rhodopsin²⁴ and for other GPCRs (see ref. 25). Notably, the H8 of the muscarinic M3R was shown to cross-link with the G_{αq} protein in an agonist-dependent manner²⁶, suggesting a close proximity between the receptor H8 domain and the G protein. Our results support a model where G protein could initially interact with the H8 and/or ICL1 domains of the receptor before docking in the 7TM core and stabilizing the TM5 and TM6 domains in an open conformation (Fig. 3c).

Here we describe the measurement of conformational changes associated with μ OR activation in both the extracellular and intracellular domains. Our data offer insights into agonist regulation of μ OR subdomain conformations, revealing a possible allosteric pathway of activation from the agonist-binding pocket to the ICL1 and H8 domains to initiate the ternary complex formation and associated changes in TM5 and TM6. Importantly, our data complement the high-resolution structural data of the μ OR agonist-induced state bound to BU72 described in a companion article (Huang *et al.*, ref. 2) and offer unprecedented insights into the structural mechanisms of opioid receptor activation. A better understanding of the structural basis for opioid receptor activation may lead to new therapeutic approaches with fewer side effects.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 March; accepted 19 June 2015.

Published online 5 August 2015.

1. Manglik, A. *et al.* Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321–326 (2012).
2. Huang, W. *et al.* Structural insights into μ -opioid receptor activation. *Nature* <http://dx.doi.org/10.1038/nature14886> (2015).

3. Manglik, A. *et al.* Structural insights into the dynamic process of β_2 -adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
4. Matthes, H. W. *et al.* Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ -opioid-receptor gene. *Nature* **383**, 819–823 (1996).
5. Macht, D. I., Herman, N. B. & Levy, C. S. A quantitative study of cutaneous analgesia produced by various opium alkaloids. *Proc. Natl Acad. Sci. USA* **1**, 582–585 (1915).
6. Pasternak, G. W. & Pan, Y. X. Mu opioids and their receptors: evolution of a concept. *Pharmacol. Rev.* **65**, 1257–1317 (2013).
7. Raehal, K. M., Schmid, C. L., Groer, C. E. & Bohn, L. M. Functional selectivity at the μ -opioid receptor: implications for understanding opioid analgesia and tolerance. *Pharmacol. Rev.* **63**, 1001–1019 (2011).
8. McPherson, J. *et al.* μ -opioid receptors: correlation of agonist efficacy for signalling with ability to activate internalization. *Mol. Pharmacol.* **78**, 756–766 (2010).
9. Stephenson, R. P. A modification of receptor theory. *Br. J. Pharmacol. Chemother.* **11**, 379–393 (1956).
10. Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
11. Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
12. Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180 (2011).
13. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504**, 101–106 (2013).
14. Nygaard, R. *et al.* The dynamic process of β_2 -adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
15. Kofuku, Y. *et al.* Efficacy of the β_2 -adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nature Commun.* **3**, 1045 (2012).
16. Kofuku, Y. *et al.* Functional dynamics of deuterated β_2 -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **53**, 13376–13379 (2014).
17. Kim, T. H. *et al.* The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J. Am. Chem. Soc.* **135**, 9465–9474 (2013).
18. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wuthrich, K. Biased signaling pathways in β_2 -adrenergic receptor characterized by ^{19}F -NMR. *Science* **335**, 1106–1110 (2012).
19. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463**, 108–112 (2010).
20. Palmer, A. G. III, Kroenke, C. D. & Loria, J. P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.* **339**, 204–238 (2001).
21. Neilan, C. L. *et al.* Characterization of the complex morphinan derivative BU72 as a high efficacy, long-lasting μ -opioid receptor agonist. *Eur. J. Pharmacol.* **499**, 107–116 (2004).
22. Knierim, B., Hofmann, K. P., Ernst, O. P. & Hubbell, W. L. Sequence of late molecular events in the activation of rhodopsin. *Proc. Natl Acad. Sci. USA* **104**, 20290–20295 (2007).
23. Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* **34**, 601–611 (2009).
24. Ernst, O. P. *et al.* Mutation of the fourth cytoplasmic loop of rhodopsin affects binding of transducin and peptides derived from the carboxyl-terminal sequences of transducin α and γ subunits. *J. Biol. Chem.* **275**, 1937–1943 (2000).
25. Kaye, R. G., Saldanha, J. W., Lu, Z. L. & Hulme, E. C. Helix 8 of the M1 muscarinic acetylcholine receptor: scanning mutagenesis delineates a G protein recognition site. *Mol. Pharmacol.* **79**, 701–709 (2011).
26. Hu, J. *et al.* Structural basis of G protein-coupled receptor-G protein interactions. *Nature Chem. Biol.* **6**, 541–548 (2010).

Acknowledgements We acknowledge support from INSERM (S.G.) and CNRS (H.D.) and from the National Institutes of Health Grant (NIDA-DA036246 to B.K.K. and S.G.). We also acknowledge the National Institute of Drug Abuse Drug Supply Program for providing [Dmt¹]DALDA.

Author Contributions R.S., H.D., A.M., W.H., B.K.K. and S.G. designed experiments, performed research and analysed data. C.M. expressed, purified and characterized receptor and nanobody preparations. J.S. and T.L. developed the G protein mimetic nanobodies. H.D. supervised NMR data analysis. S.G. and R.S. prepared the manuscript with the help of H.D. and B.K.K. S.G. supervised the overall project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G. (sebastien.granier@igf.cnrs.fr) or H.D. (helene@cbs.cnrs.fr).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification. Nanobody (Nb) sequences were subcloned into a pMalp2x vector containing an N-terminal, 3C protease-cleavable maltose binding protein (MBP) tag and a C-terminal 8 × His tag. Plasmids were transformed into BL21(DE3) cells and protein expression induced in lysogeny broth (LB) by addition of IPTG to 1 mM at an OD₆₀₀ of 0.8. After 24 h of incubation at 20 °C, cells were collected and sonicated in the presence of lysozyme and benzonase. MBP–nanobody fusions were purified by Ni-NTA chromatography and MBP was removed using 3C protease. Cleaved MBP was separated from the nanobody by additional amylose purification and size exclusion chromatography in a buffer containing 20 mM HEPES pH 7.5 and 0.1 M NaCl. Of note, because we needed large amounts of Nb for the NMR experiments, we decided to use Nb33 throughout this study as it was produced and purified in much larger quantities than Nb39 (almost ten times as much for the same volume of BL21(DE3) cell culture).

We generated a μOR mouse construct with features designed to enhance stability for NMR spectroscopy. A tobacco etch virus (TEV) protease recognition site was introduced after residue 51, and a human rhinovirus 3C protease site after residue 358. A Flag epitope tag was added to the amino terminus and an 8 × His tag was appended to the carboxy terminus. Receptor expression was largely improved by using a M72T single-point mutation as assessed by naloxone binding. The final construct (μOR–2x) is shown in Extended Data Fig. 1.

The μOR–2x construct was expressed in Sf9 cells using the Bestbac (Expression systems) baculovirus system in the presence of 3 μM naloxone. For all other mutants (assignments), we used the bac-to-bac system (Life Technologies). Sf9 cell cultures (Life Technologies) were grown to a density of 4×10^6 cells per ml, infected with baculovirus containing the μOR–2x genes, shaken at 27 °C for 48 h, and cell pellets were harvested and stored at –80 °C. The receptor was purified as previously described¹ and eluted from the anti-Flag M1 affinity resin with a buffer containing 0.01% MNG, 0.001% CHS, 20 mM HEPES pH 7.5, 0.1 M NaCl, 1 μM naloxone, 0.2 mg ml^{–1} Flag peptide and 2 mM EDTA.

To remove flexible amino and carboxy termini, TEV and 3C protease were added at a 1:5 and 1:10 protease:μOR–2x ratio by weight. The sample was incubated at 4 °C overnight in the presence of 100 μM of TCEP. We then used a negative Ni-NTA chromatography step to remove TEV and 3C proteases.

μOR reductive methylation. Receptor preparation from the Ni-NTA flow through were incubated at 4 °C overnight with 10 mM ¹³C-formaldehyde and 10 mM NaBH₃CN. Excess of reagent was eliminated by dialysis and ¹³CμOR was further purified by SEC chromatography in a buffer containing 0.01% MNG, 0.001% CHS, 20 mM HEPES pH 7.5 and 0.1 M NaCl.

The monodisperse peak was then concentrated to 20 to 40 μM final, and dialysed in 98.85% D₂O buffer with 0.01% MNG, 0.001% CHS, 20 mM HEPES-d18 pH 7.4 (uncorrected) and 40 mM NaCl.

NMR spectroscopy. Final samples (~270 μl at 20–40 μM) were loaded into Shigemitsu microtubes susceptibility matched to D₂O. All data for ligands and mutant studies were acquired on 700 MHz and 500 MHz Bruker Avance III spectrometers, respectively (Bruker, Rheinstetten, Germany), both equipped with

5 mm cryogenic H/C/N/D probes with z axis gradient. ¹H–¹³C correlation spectra were recorded using heteronuclear multiple-quantum coherence (HMQC) experiments in echo/anti-echo mode. ¹³C and ¹H chemical shifts and peak line widths in the HMQC spectra reveal the chemical and magnetic environments of the ¹³C-methyl probes as well as their dynamic properties. Spectral widths in ω₁ and ω₂ were 8,417.5 Hz (6,009.6 Hz) and 3,519.6 Hz (2,515.2 Hz) at 700 MHz (500 MHz), respectively centred at 40 p.p.m. in the ¹³C dimension. ¹³C decoupling was performed with a GARP4 sequence. Typically, 134 (81) complex points with 32–48 scans per FID were recorded, to ensure a 27-Hz resolution per point at 700 MHz (500 MHz) before zero filling. The relaxation delay was set to 1.5 s. Thirty-two steady-state scans preceded data acquisition. Total collection time varied between 3 and 4 h, depending on the sample concentration. Spectra were processed using NMRPipe/NMRDraw software²⁷ and visualized using CCPNMR²⁸.

All ligands were dissolved in deuterated DMSO to 10 mM and directly added to the sample in the Shigemitsu tube at a final concentration of 150 μM. Nb33 and Nb39 were concentrated to 0.6 mM and dialysed in 100% D₂O buffer with 0.01% MNG, 0.001% CHS, 20 mM HEPES-d18 pH 7.4 (uncorrected) and 40 mM NaCl. The nanobodies were added directly in the Shigemitsu tubes at a final concentration of 60 μM before data acquisition.

Peak volume analysis. All NMR spectra were processed using the suite of programs provided in the NMRPipe/NMRDraw software distribution²⁷. The spectra were normalized using DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) as an internal reference. For peak fitting analysis, spectra were processed with a squared sine-bell window function in each dimension, and zero-filled to 4096 × 1024 data points in time domain data *t*₂ and *t*₁, respectively. Spectra were fit with the program nlinLS, provided as a part of nmrDraw package. Gaussian models were used for the fitting in each dimension, starting from values obtained from the peak-picking routine in nmrDraw. The quality of the fits was examined visually by estimating the residual difference between the experimental data and the results of the model calculations. Peak volumes in apo-state, ligands alone (BU72, [Dmt¹]DALDA, and DAMGO) and ternary complexes spectra were extracted from the peak fitting. Errors in the peak volume were calculated based on the effect of random noise for the peak height estimated by nlinLS.

To directly compare the ligand effects on the apo-state relative to the maximal effect obtained with BU72–Nb33 treatment (fully agonist-induced state), we normalized the peak volumes with the volume difference between the apo-state and ternary complex condition (BU72–Nb33) as the 100% (*V*_{apo} – *V*_{BU72–Nb33}). Ligand effects are represented as percentage of the maximal effect:

$$100 \times \frac{V_{\text{apo}} - V_{\text{L or L-Nb33}}}{V_{\text{apo}} - V_{\text{BU72-Nb33}}}$$

For the TM6 data, because of the overlapping peaks in the area of interest, we considered the peak volume of K269^{6,24} and K271^{6,26} in the ternary complex with BU72 as the 100% (indicated as K269* and K271* in

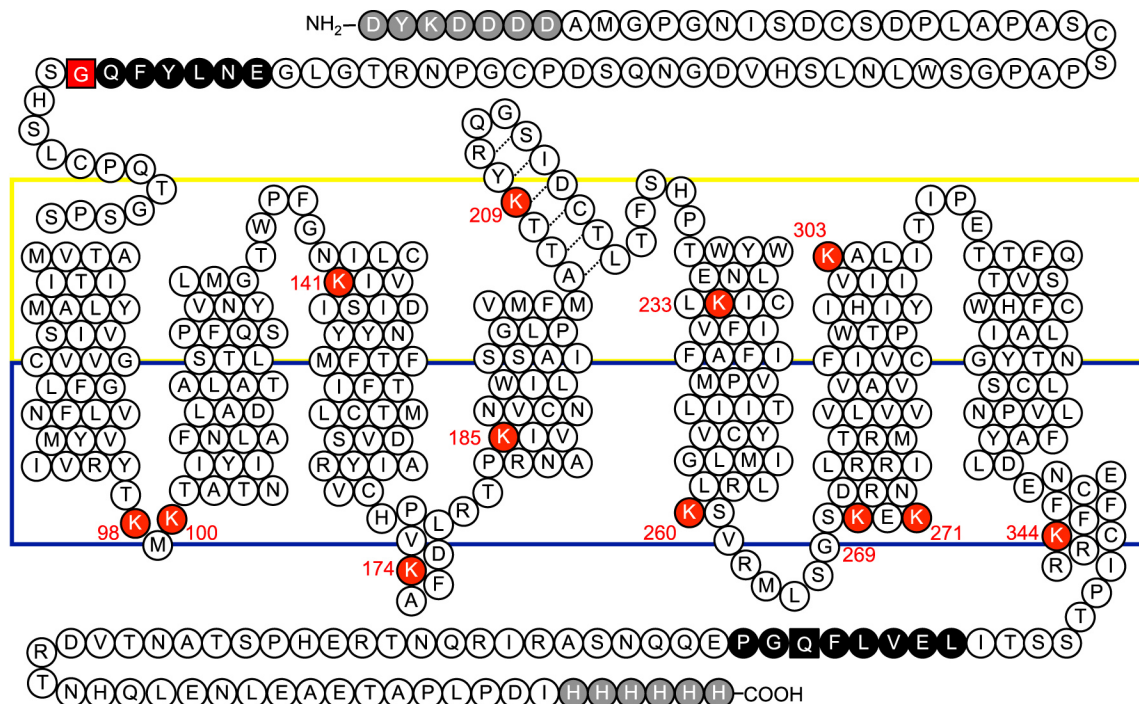
Fig. 2) and in this case the percent of maximal effect is $100 \times \frac{V_{\text{L or L-Nb33}}}{V_{\text{BU72-Nb33}}}$. The

uncertainty was determined by $\sqrt{\Delta V_{\text{L or L-Nb33}}^2 + \Delta V_{\text{apo}}^2 + \Delta V_{\text{BU72-Nb33}}^2}$.

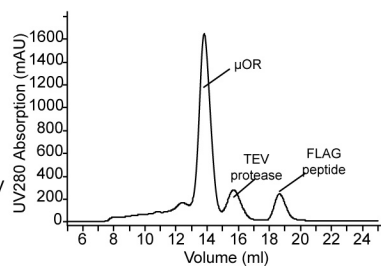
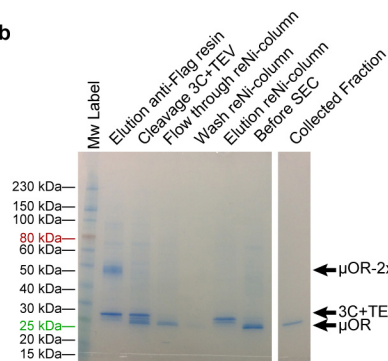
27. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

28. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).

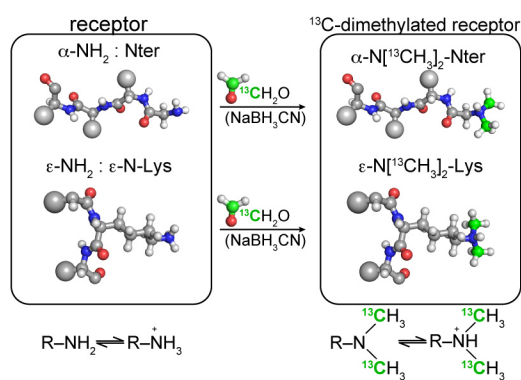
a



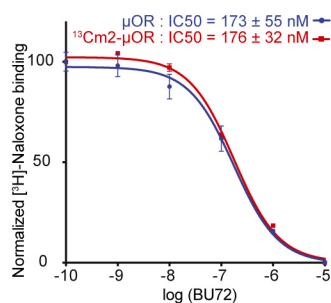
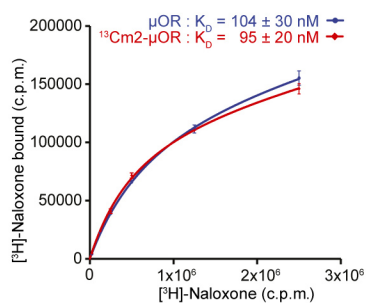
b



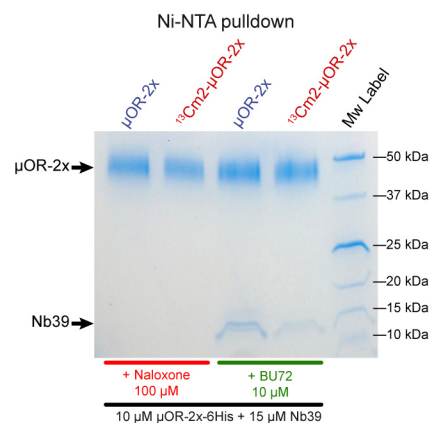
C



d

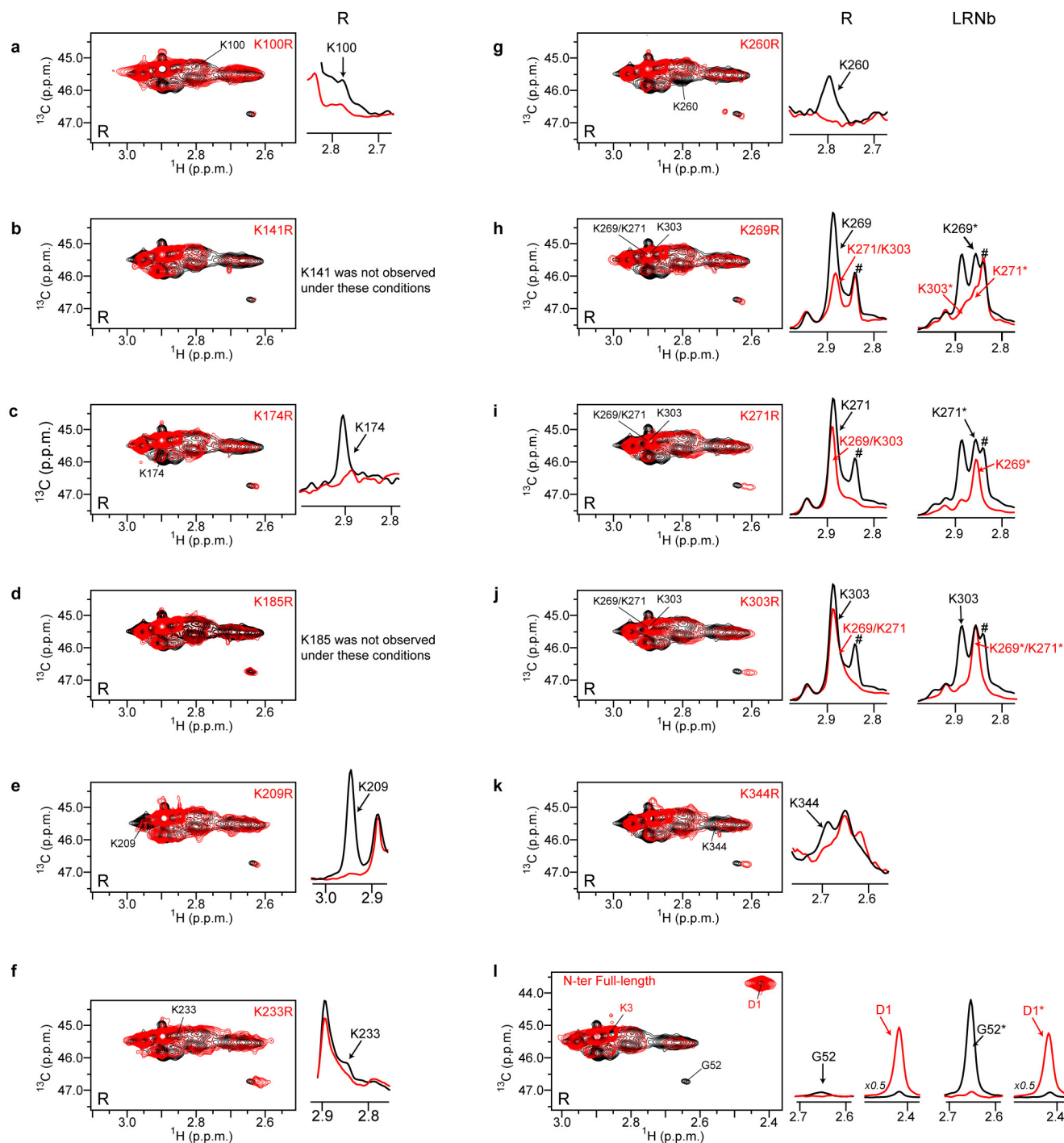


e



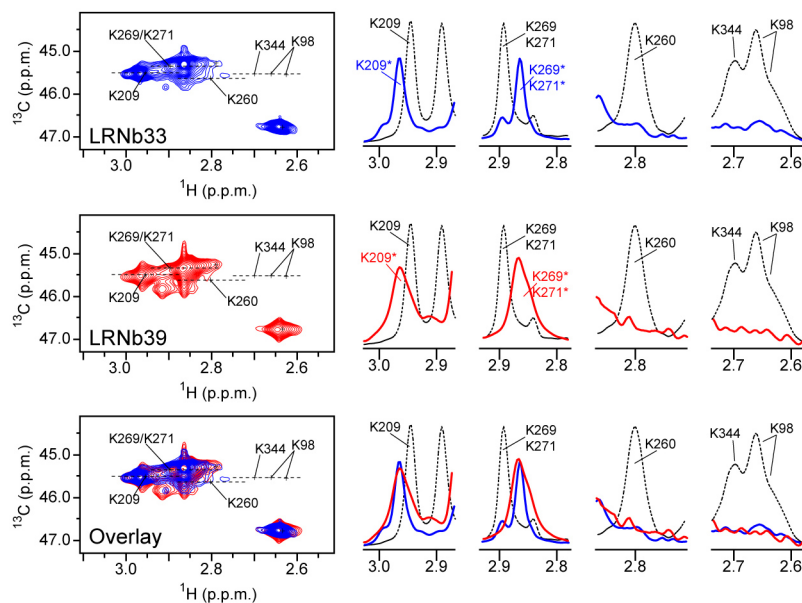
Extended Data Figure 1 | Overall strategy for the preparation of ^{13}C -dimethylated μOR and quality control. The wild-type μOR contains a total of 12 lysine residues strategically positioned to sense conformational dynamics in both extracellular and G-protein-coupling domains upon receptor activation. To monitor the μOR activation process in solution in an amphiphile membrane-like environment, we exploit the sensitivity of two ^{13}C -methyl groups covalently bound to the $\epsilon\text{-NH}_2$ of lysine side chains ($\epsilon\text{-N}[^{13}\text{CH}_3]_2\text{-lysines}$) and to the $\alpha\text{-NH}_2$ of the receptor N terminus as NMR probes, an approach recently validated in the GPCR field¹⁹. We slightly modified the wild-type sequence to facilitate receptor purification and to remove flexible N and C termini for improvement of sample stability in NMR measurements. **a**, Snake plot presenting the μOR -2x construct and highlighting the 12 endogenous lysines (red circles), the protease cleavable motifs (black circles) and the Flag and $6 \times$ histidine tags (grey circle). Extracellular and intracellular surfaces are coloured yellow and blue, respectively. **b**, The different biochemical steps were analysed by SDS-PAGE (left panel) and a typical final size-exclusion chromatography highlighting the monodispersity of ^{13}C -dimethylated μOR is shown (right panel). **c**, Shown is the reductive methylation of the N terminus and of lysine side chains after reaction with ^{13}C -formaldehyde (^{13}C in green) in the presence of sodium cyanoborohydride (NaBH_3CN). Amino acids are shown in ball-and-stick representation and, for clarity, side-chains of unmodified amino acids are replaced by large spheres.

This reaction is known to minimally affect protein structure and function and, as previously observed with $\beta_2\text{-AR}$ (see ref. 19), it did not affect the functionality of μOR as ^{13}C -dimethylated μOR binds both antagonist (Naloxone) and agonist (BU72) with a similar affinity than unlabelled μOR (**d**). The left panel in **d** represents the mean \pm s.d. of triplicates for one experiment representative of three experiments in total ($n = 3$). The right panel in **d** represents normalized values \pm s.d. (larger binding value being 100% and smaller 0%) for one experiment representative of three experiments in total ($n = 3$). K_d and IC_{50} values are the mean \pm s.d. of the three independent experiments for both receptor preparations. In addition, we observed similar agonist-induced interaction with Nb39 between μOR or ^{13}C -dimethylated μOR using pull-down experiments (**e**), demonstrating that methylation does not prevent the agonist-dependent interaction with Nb39. Saturation binding experiments on soluble μOR or ^{13}C -dimethylated μOR (both at 100 nM) were done in the presence of an increasing amount of radiolabelled naloxone (up to 1 μM) and non-specific binding was determined in the presence of 100 μM of naloxone. Competition assays on soluble μOR or ^{13}C -dimethylated μOR were done with 100 nM of radiolabelled naloxone and increasing amount of BU72. Free and receptor-bound radioligands were separated using gel-filtration columns. Total binding was plotted as a function of [^3H]-naloxone or BU72 concentration and data were analysed using Prism with saturation or competitive binding analyses.



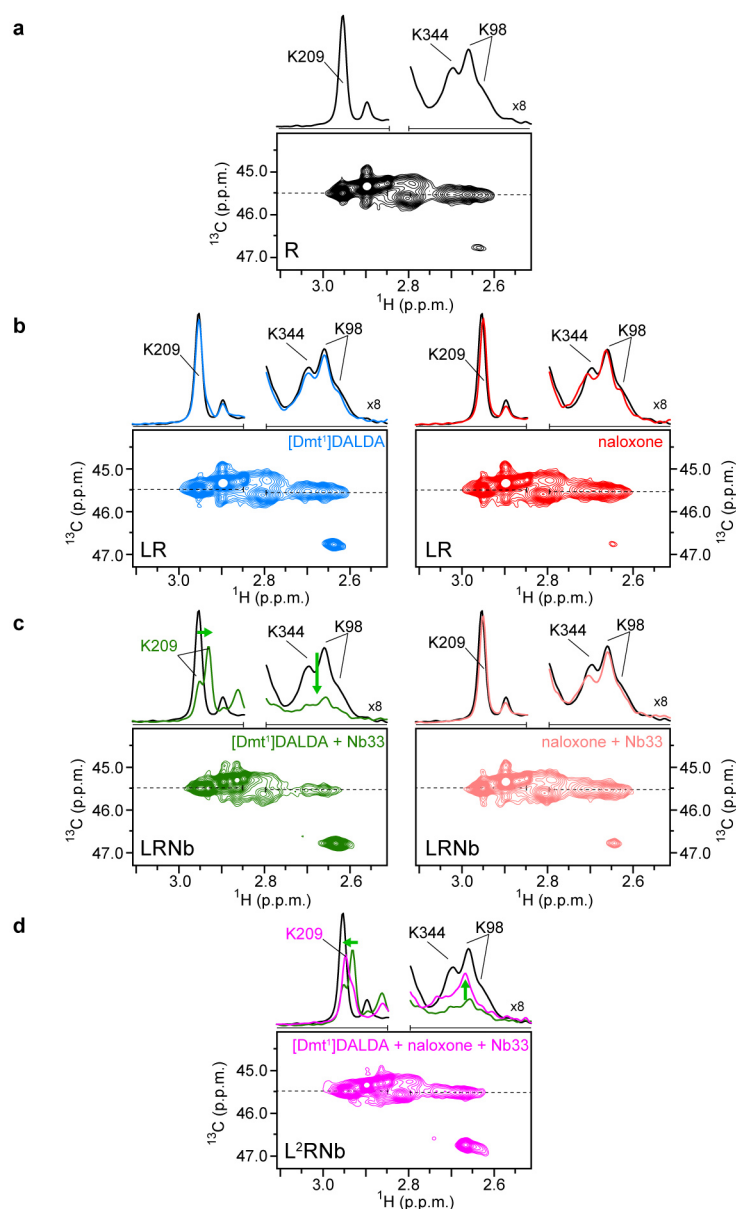
Extended Data Figure 2 | Assignments of the N terminus and dimethyllysine peaks. Each single mutant was expressed, purified and labelled as described for $\mu\text{OR}-2x$. For each mutant, we analysed the R and LRNb spectra. **a–k**, Left panels represent the 2D spectra with the indicated lysine to arginine mutant in red. On the right panels, we highlight the peak disappearance for each mutant (red line) as compared to the wild-type μOR preparation (black line) in the ^1H dimension in the receptor alone condition (R). In some instances, peaks are overlapping in the area of interest and to assign these peaks we had to record the NMR spectra in the ternary complex situation (LRNb). This was the case for K269 (**h**), K271 (**i**), and K303 (**j**). In **h**, **i** and **j**, # indicates the position of the lysine peak from the Flag tag that we measured in some samples and that is due to an incomplete cleavage of the N terminus by the 3C protein. For samples used in the analysis of ligand and Nb effects, we almost eliminated this peak by optimizing the 3C cleavage step as described in the

Methods (treatment with 100 μM TCEP). The peak denoted with a # in panel **i** between wild type (no TCEP) and K271R (100 μM TCEP) is an example of the improvement of 3C cleavage in TCEP-treated samples as indicated by the almost complete disappearance of the peak in both R and LRNb conditions (K271R, red line). The K98 peak was assigned by deduction (Fig. 1a). Indeed, we generated and recorded NMR spectra of 11 mutants of the 12 endogenous lysines and the only peaks that were never affected in all the mutant spectra and left unassigned necessarily correspond to the K98 residue. **l**, The assignments of D1 (full-length N terminus) and G52 (cleaved N terminus) were inferred from the spectra obtained before (red line) and after cleavage with 3C (black line). The G52 loss of signal is better observed in the LRNb situation (G52* almost completely disappears). Asterisks indicate the chemical shifts of methyl probes in the ternary complex condition.



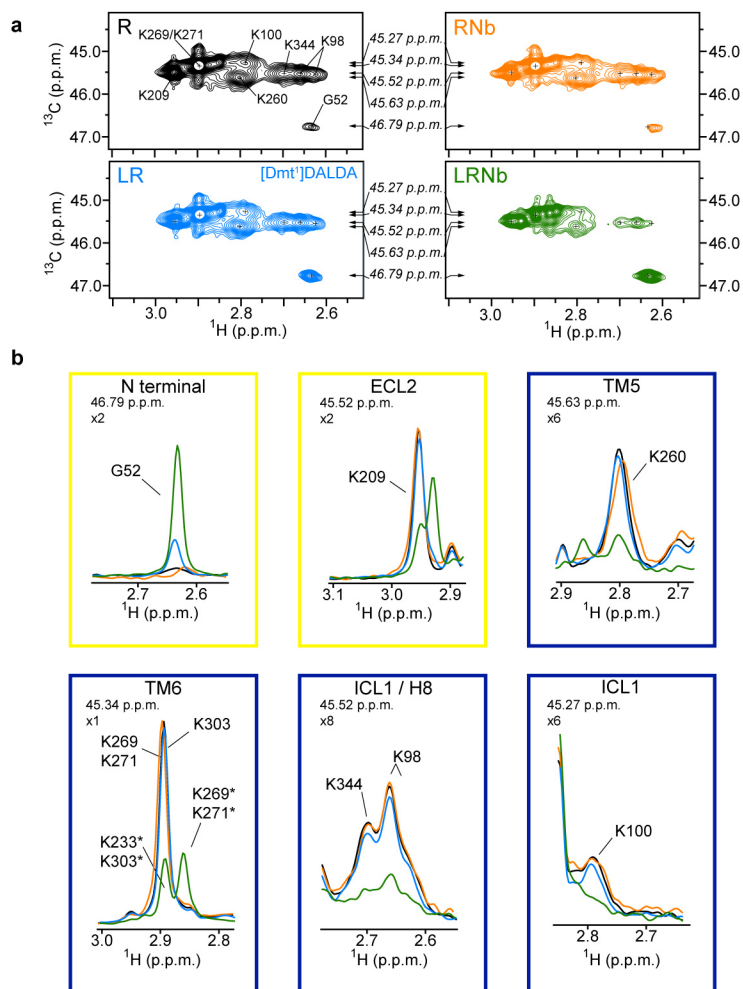
Extended Data Figure 3 | Comparison of Nb33 and Nb39 effects on NMR spectra. HMQC spectrum of ^{13}C -dimethylated μOR in the ternary complex situation (R, 30 μM ; BU72, 150 μM ; Nb, 60 μM) for Nb33 acquired with a 700 MHz spectrometer (blue trace) and for Nb39 acquired with a 500 MHz spectrometer (red trace). Both spectra are very similar and are characteristic to

the fully agonist-induced state of μOR . 1D slice of HMQC spectra in the ^1H dimension (right panels) highlighting the similar effect of Nb33 and Nb39 in the presence of BU72 on the indicated set of dimethyllysines and compared to the untreated sample (black dotted line).



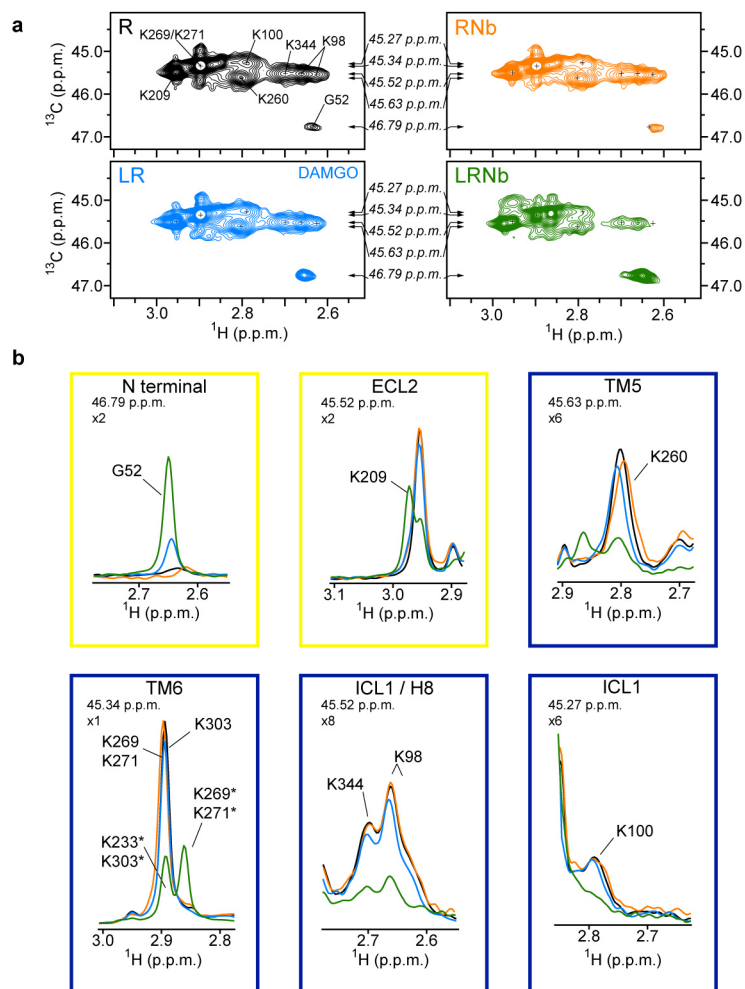
Extended Data Figure 4 | Partial reversal of agonist effects by treatment with the antagonist naloxone. a–c, HMQC spectra of unliganded ^{13}C -dimethylated μOR (30 μM , R, black) (a), the same sample bound to a saturating concentration of ligand (150 μM , LR, [Dmt 1]-DALDA (blue) or naloxone (red) (b) and the same sample treated with a saturating concentration of Nb33

(60 μM , LRNb, [Dmt 1]DALDA (green) or naloxone (light pink) (c). d, Spectra of the [Dmt 1]DALDA–RNb sample treated with naloxone (900 μM) (L 2 RNb, magenta). Green arrows in 1D traces highlight the peaks movement in the activation process (c) and for the partial reversal effect of naloxone (d) for K209 and K344/K98 peaks.



Extended Data Figure 5 | Activation of μ OR by [Dmt¹]DALDA. **a**, HMQC spectra of unliganded ^{13}C -dimethylated μ OR (30 μM , R, black), ^{13}C -dimethylated μ OR treated with a saturating concentration of [Dmt¹]DALDA (150 μM , LR, blue), ^{13}C -dimethylated μ OR treated with [Dmt¹]DALDA and a saturating concentration of Nb33 (60 μM , LRNb, green) and of unliganded ^{13}C -dimethylated μ OR treated with a saturating concentration of Nb33 (60 μM , RNb, orange). The small crosses indicate the position of peaks visible in the apo-state that are reduced in intensity or no longer visible in the other spectra.

Arrows indicate the position of the 1D slice represented in **b**. **b**, 1D slice of HMQC spectra in the ^1H dimension highlighting the effect of [Dmt¹]DALDA (blue) and [Dmt¹]DALDA + Nb33 (green) on the N terminus (G52) and on the indicated set of dimethyllysines (ECL2, TM5, TM6, ICL1 and H8 lysines). The yellow and blue boxes indicate the positions in the extracellular and intracellular domains respectively. The zoom level for each peak and the ^{13}C chemical shifts are indicated in the top left of each panel. Asterisks indicate the chemical shifts of methyl probes in the ternary complex condition.



Extended Data Figure 6 | Activation of μ OR by DAMGO. **a**, HMQC spectra of unliganded ^{13}C -dimethylated μ OR (30 μM , R, black), ^{13}C -dimethylated μ OR treated with a saturating concentration of DAMGO (150 μM , LR, blue), ^{13}C -dimethylated μ OR treated with DAMGO and a saturating concentration of Nb33 (60 μM , LRNb, green) and of unliganded ^{13}C -dimethylated μ OR treated with a saturating concentration of Nb33 (60 μM , RNb, orange). The small crosses indicate the position of peaks visible in the apo-state that are reduced in intensity or no longer visible in the other spectra. Arrows indicate the

position of the 1D slice represented in **b**. **b**, 1D slice of HMQC spectra in the ^1H dimension highlighting the effect of DAMGO alone (blue) and DAMGO + Nb33 (green) on the N terminus (G52) and on indicated set of dimethyllysines (ECL2, TM5, TM6, ICL1 and H8 lysines). The yellow and blue boxes indicate the positions in the extracellular and intracellular domains respectively. The zoom level for each peak and the ^{13}C chemical shifts are indicated in the top left of each panel. Asterisks indicate the chemical shifts of methyl probes in the ternary complex condition.

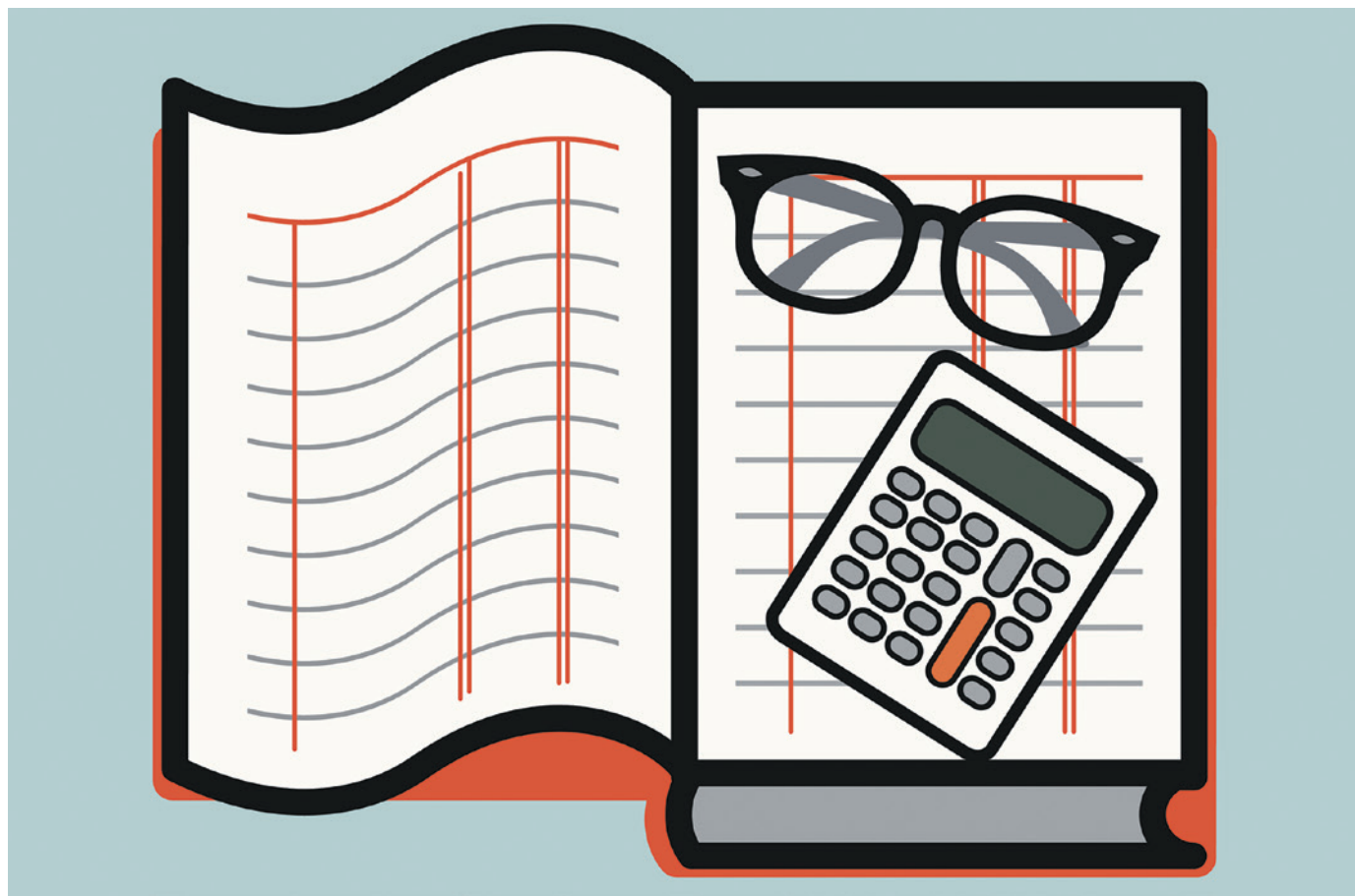
CAREERS

PLETHORA OF PATHS From postdoc to press and industry go.nature.com/evaidb

WRITING COMPETITION Publish better science through better data go.nature.com/zsx1is

NATUREJOBS For the latest career listings and advice www.naturejobs.com

LUCIANO LOZANO/GETTY



COLUMN

The dream budget

The key to a strong grant application is a well-considered and detailed outline of expenses, says grant director **Ingrid Eisenstadter**.

Several times a year, my inbox fills with grant proposals from scientists who are seeking support for their work. I am the director of grants for a small family foundation that has supported research for more than half a century. My desk is the first step in a process that will determine the fate of these funding requests. I start by reading the proposal's title and the introductory summary of its significance.

Then, I turn to the budget page.

The first thing I look at is the bottom line. If the full request is in our range, I skim through the rest with an eye for potential red flags. For example, we have a ceiling on overhead requests: we do not support operating funds or proposals that are primarily for equipment. Only after I have assessed whether there are issues with the budget do I return to the proposal itself.

When drafting a budget, it is crucial to know the maximum grant amount awarded by the foundation to which you are applying. If you

ask for much more, the granter might reject your proposal on that basis alone. Some organizations avoid publishing an exact figure to discourage applicants from reflexively requesting it. How can you find out? In the United States, the Foundation Center, a national database for grant-seekers, provides access to the 'Form 990' tax return that all foundations are required to submit to the federal government (see go.nature.com/panexz). These returns disclose the dollar range of grants issued by a foundation each ►

► year. You should also look at the granters' websites. Most do not post their 990 forms, but some have annual reports or web pages that list project titles or recipients and the corresponding grants' dollar amounts. Be aware that when granters see budgets, for example, for exactly US\$25,000 or \$50,000, many will suspect that aspects of the budget were contrived to reach this exact amount. Proceed with caution.

It is also important to find out whether a foundation has specific requirements for the budget section of grant applications — and, if so, to observe them strictly. Many do, and they typically post this information on their websites along with their application guidelines. These rules will tell you whether you can, for instance, purchase equipment, support undergraduate assistants or request publication fees for journal articles that may or may not arise from your work. They also specify whether there is a ceiling for overhead — or 'indirect' — expenses. With our limited resources as granters, we want to fund research, not depreciation and administration, so if an institution submits an application that proposes adding 50–70% — or even more — to the bottom line for overhead expenses, that can be a problem.

The budgeting rules will also tell you whether you can ask for general operating support or only specific project-related funds, and whether the funder will award multiple-year grants. Find out before you start budgeting. If your questions are not answered on the website, send an e-mail to the foundation to ask for clarification. There

is nothing unusual about these enquiries.

The majority of the budget requests that we receive are indeed for our maximum grant amount, and many staff members from other grant-giving foundations have told me the same thing. Or, put another way, they have the same complaint. The place where we tend most often to see this escalation is in the budget's salary line.

SALARY SAMPLES

We received an e-mail proposal recently that requested US\$14,550, with more than half of that allocated to salary. When the hard copy of that proposal arrived, however, it looked as though the applicant had accessed our 990 form, because the salary allocation — in an otherwise unchanged proposal — had more than doubled. The applicant had added months' worth of support, raising the total request to \$24,550.

We also received a proposal in which the principal investigator indicated that he had asked a second funder for \$15,000, in addition to the \$25,000 that he was seeking from us. I asked to see that second budget and discovered that he would actually require an extra \$1,700 if he received only the smaller grant. That is, a total of only \$16,700 would have allowed this preliminary investigation to proceed — in the absence of any salary support at all. He would have had to scale back his project, but its successful completion would have nonetheless enabled him to make a larger request for state or federal funding.

Although both of these researchers should have explained their position, I did not consider either case to be an act of duplicity (we funded one of them). An applicant's use of the salary line to increase a grant request is understandable to — if not welcomed by — the granter. There is prestige for a researcher in obtaining large grants, and some institutions do require staff to bring in salary support.

You should visit your development office early in the budgeting process to talk to staff members and learn what they expect. Most universities and non-governmental organiza-

"If a grant application has sparkle — novelty and a thoughtful protocol — it could persuade us."

tions are prepared to reduce their overhead requirements to adhere to foundations' guidelines. We often see budgets in which the home institution produces matching funds to enable its researchers to stay within our

support guidelines. If this is the case, consider adding that information to your budget page in a column labelled 'matching funds' or 'other sources' to show that your request is backed by a contribution to salary, fringe benefits, overhead expenses or any other support your home institution is willing to provide.

You might also want to include a column for other sources when you are applying to two or more granters for the same project. Here, it is important to explain in your budget justification whether the funds from these sources are already in hand (see 'Explain your budget'). If not, you need to clarify whether the project will still proceed — perhaps on a smaller scale — if the other organizations fail to contribute. State in your application when you will know about any other funds. The timing is important because a granter will be reluctant to write you a cheque if your project runs the risk of halting midway for lack of that support.

Too much information in a budget is better than too little. On my desk recently was a proposal from a principal investigator who wanted to travel to South America — with her dog. This added, yes, 13 lines to her budget, ranging from \$30 for a flea-and-tick collar to \$400 for a canine plane ticket. The researcher explained that her dog is not merely her travel companion but also a well-trained scat-detection dog — a necessity for her research. We funded this interspecies team.

If a grant application has sparkle — novelty and a thoughtful protocol — it could persuade us to work with the researcher to reduce the final request to better fit our resources. Or we may even agree to write a cheque that is larger than usual. ■

Ingrid Eisenstadter is director of grants for The Eppley Foundation for Research in New York City.

WALK THE LINE

Explain your budget

Although not all granters specify that they want to see a page entitled 'Budget justification', it is wise to include one. This is a line-by-line explanation of itemized expenses in your request. Usually, just a few sentences are sufficient to address each expenditure. Also, keep in mind that your proposal could be going to a foundation whose evaluators are not trained in your speciality. To the extent possible, your budget justification should be written in language that is clear to a layperson.

A good place to start is the salary line. We often see a one- or two-month salary request for projects that will last a year. Explain this. For instance, one common reason for the apparent discrepancy is that principal investigators devote just part of their time during the year to the proposed projects, so the request would cover 100% of salary support for the work. In other instances, home institutions might pay staff researchers' salaries during the academic year and expect them to raise salary funds only for the summer months.

Whatever the reason, you need to clarify. If you are asking for a salary or stipends for undergraduate assistants or PhD students, include in your budget an hourly or daily rate or the percentage of time that they will devote to the project, so that granters understand the exact time commitment that they are being asked to fund.

If you need to pay for supplies or tests, name them and explain how they will be used. The same goes for an equipment line. Explain why you need a new incubator or how you will use the desiccator, centrifuge or imaging equipment. If you are purchasing equipment that will outlive your research project, consider briefly explaining how you will use it in the future, so that the foundation understands the lasting benefits of its long-term investment. If you are asking for food and lodging support for travel, list a daily expense that explains the total.

When you neglect to provide these explanations, you are asking evaluators to fish through your text for details about each budget line. This is not to your advantage. **I.E.**

IN A NEW LIGHT

Career opportunities.

BY S. R. ALGERNON

“O Lucifer,” said Ned Goodman, sotto voce, as he knelt at a stone altar in a clearing far from town. “Take corporeal form. Appear before me. Offer me your power and I will do as you command.”

Goodman placed his hands on an iron-bound codex that rested on the altar. *Soon*, he thought, *there will be no more weekends running focus groups for toothpaste and deodorant.* Soon, he would wield real power.

Goodman had studied every known ancient language, but the markings on the pages meant nothing to him. The phonetic transcription — on foolscap and weighed down with a rock — had come from a deranged monk of a disgraced order. He had said — among his more lucid ravings — that Hephaestus had forged the codex aeons ago.

Goodman read the incantation. As he did, the earth shook, and the ground on the other side of the altar collapsed into a molten pool. Stray tufts of grass near its edge caught fire. The air stank of sulfur, and it stung Goodman’s eyes. A brilliant column of liquid metal oozed up from the pool, taking on a red glow as it cooled. It ballooned out on top like a mushroom or a jellyfish. Flames erupted from its surface, as if the protuberance wore a corona or a ring of horns.

“Are you Lucifer?” asked Goodman.

“You might say I am a servant thereof.” The glowing form grinned, revealing a tongue of flame behind pointed teeth. It extended a hand, sheathed in flame.

“Follow me. Lucifer awaits.”

“Follow you? Where?”

“To a world of fire. You shall bathe in acid under the weight of a stormy, sulfurous sky.”

Goodman shrank back.

“I know you serve the Prince of Darkness and all,” said Goodman, “and I don’t mean to be impertinent, but aren’t you supposed to grant me fearsome powers before dragging me off to perdition?”

“Powers?” The agent of Lucifer’s smile faded. “We have geothermal power, if that’s what you mean.”

“Aren’t I supposed to get a chance to do some sinning first?”

“Are you that monk that’s been bothering us? Look. We told you last time...”

Discordant howls from beyond the portal drowned out the rest of the reply. A spray of shimmering metal drenched the agent from the right side, dulling the glow of its horns



momentarily. A few stray droplets landed on the altar, setting the foolscap alight.

“Hey, you two! No splashing...! Where was I? Look, I don’t have time to argue with you. Did you request a portal to Lucifer or not?”

“I did.”

“Lucifer? The Morning Star?”

“I...”

“The Evening Star?”

“Is that a tabloid or something?”

“Venus?”

“Are you telling me that that molten glob and those choking fumes are from Venus?”

“What else would they be?”

What good was Venus? thought Goodman. He recalled the people who had mocked his quest over the years and considered throwing them into the portal. That would certainly be devilish, but he couldn’t quite muster a thirst for vengeance. He had been so busy tracking down the codex that he hadn’t had time to acquire any real enemies.

“But what about the contract signed in blood?” asked Goodman. “What about the pitchforks and tormented souls?”

“I... don’t really...” said the agent of Lucifer. “Could you hold for just a minute?”

Goodman watched the ground smoulder, wondering if he should just take the codex and go home. An iron beach ball bounced by within the swirling haze. Vague shapes galloped after it.

“So sorry to keep

you waiting,” said the

agent of Lucifer upon

its return. “We closed

down our Earth operations centuries ago, including the guest-worker programme.” The agent glanced down at the codex. “It seems that your information is out of date.”

Goodman sighed and then coughed as a stray plume of carbon dioxide extinguished a smouldering corner of the notes.

“You’re saying you don’t want my soul?”

“I’m afraid not. Look, I know we should have been clearer about the policy changes, and I hate to see you leave empty-handed.”

The agent’s eyes brightened to a pleasant shade of orange. “I know! Some of our historical Earth clients have requested heavy metals. Our factories had a cancelled order from the Ishtar Terra Children’s Marching Band. Could I interest you in an oboe made of osmium?”

“No, I don’t think...”

“A tuba made of tungsten?”

“No. You really don’t want my soul? I figured souls would still be marketable in this day and age. There’s so much evil in the world.”

“Times change. Back then, we needed someone to help us bring organics and the like back from Earth. Ever since the Oxygen Catastrophe, we’d been short-changed on raw materials, but then you came along, with your city-states and language capacity. We figured rather than letting you waste your potential hitting each other with swords... Anyway, now that we have the aerostats, we get what we need from comets and asteroids.”

“Aerostat...?”

“Buoyant craft, about a mile wide. We use them as construction platforms and launch sites.”

“You have cloud cities?”

“They’re automated, except for the maintenance crew — descendants of our original human workforce. It’s far too cold up there for our liking. There’s not much up there, now that the new resort is up and running. Probably just pools of liquid water and some overgrown greenhouses, I imagine. I could offer you a job up there, but you’d probably find it boring.”

“I was thinking something more along the lines of interplanetary tourism director. If you’re going to start up your Earth operations again, you need to rethink your image.”

“We do?”

Goodman thumped the codex for emphasis. “Absolutely. First, see if your factories can whip up a big ol’ pair of pearly gates.” ■

S. R. Algernon studied fiction writing and biology, among other things, at the University of North Carolina at Chapel Hill. He currently lives in Singapore.

ILLUSTRATION BY JACEY